



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 114014

**TO: Michael Borin
Location: REM/2A55
Art Unit: 1631
Wednesday, February 11, 2004**

Case Serial Number: 09/888615

**From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203**

edward.hart@uspto.gov

Search Notes

Examiner Borin,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

114014

From: Borin, Michael
Sent: Tuesday, February 10, 2004 4:17 PM
To: STIC-Biotech/ChemLib
Subject: Search request:09/888615; protease

Examiner: M.Borin
AU: 1631

Remsen 2A55
Tel.: 20713

RECEIVED
FEB 10 2004
STIC/CHEM. LIB.
(STIC)

RE: 09/888615; protease

Please conduct search of polypeptide SEQ ID 73 against the commercial and interference protein databases.

Thank you

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: *2/11/04*
Date Completed: *2/18/04*
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: *1*
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: *058*
WWW/Internet: _____
Other (specify): _____



STIC SEARCH RESULT FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches. Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rapn. Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn.

Because they contain data that is confidential, the results of Pending database searches should not be left in the case.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:35:56 ; Search time 52 Seconds
(without alignments)

7959.930 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKSRIGFLSYEALRRV.....DTSSMDDEDFSYKVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8563	100.0	1604	Q8NFA0	Q8nfa0 homo sapien
2	6807	99.5	1274	Q8EX85	Q8bx85 homo sapien
3	4344	53.1	1406	Q8IWI4	Q8iwt4 homo sapien
4	4440	51.9	828	Q9Y591	Q9y591 homo sapien
5	2914	34.0	1736	Q9VW49	Q9vw49 drosophila
6	2604	30.4	492	Q8N3C6	Q8n3c6 homo sapien
7	1668	19.5	354	Q95482	Q95482 homo sapien
8	1406	16.4	786	Q15634	Q15634 homo sapien
9	1233	14.4	1175	Q8WT43	Q8wt43 caenorhabdi
10	1223.5	14.4	1178	Q8WT44	Q8wt44 caenorhabdi
11	1102	12.9	952	11 Q9R085	Q9r085 rattus norv
12	1095.5	12.8	915	11 Q921M8	Q921m8 mus musculu
13	1092	12.8	973	13 Q8AVB6	Q8avb6 xenopus lae
14	1086	12.7	962	11 Q8BTL9	Q8btl9 mus musculu
15	1063	12.4	950	13 Q9PWC6	Q9pwc6 gallus gall
16	1053.5	12.3	921	4 Q8IUG6	Q8iug6 homo sapien

17	1018.5	11.9	910	10 Q93Y01	Q93y01 arabidopsis
18	1006	11.7	937	10 Q9T0B8	Q9t0b8 arabidopsis
19	1004.5	11.7	1028	10 Q9ZS85	Q9zsb5 arabidopsis
20	981.5	11.5	887	10 Q9MAQ3	Q9maq3 arabidopsis
21	974.5	11.4	924	10 Q22207	Q22207 arabidopsis
22	967.5	11.3	928	10 Q9T0B6	Q9t0b6 arabidopsis
23	939.5	11.0	887	10 Q9ZS86	Q9zsb6 arabidopsis
24	902.5	10.5	699	11 Q99K46	Q99k46 mus musculu
25	901.5	10.5	699	11 Q8C0C2	Q8c0c2 mus musculu
26	875.5	10.2	901	10 Q9C585	Q9c585 arabidopsis
27	836	9.8	951	10 Q8H557	Q8h557 oryza sativ
28	787.5	9.2	890	10 Q8LMT7	Q8lmt7 oryza sativ
29	756.5	8.8	900	5 Q9TYH8	Q9tyh8 caenorhabdi
30	675	7.9	1095	5 Q19132	Q19132 caenorhabdi
31	651	7.6	366	11 Q8BX45	Q8bx45 mus musculu
32	607	7.1	1272	5 Q9W117	Q9w117 drosophila
33	605.5	7.1	1332	5 Q8I3U1	Q8i3u1 plasmodium
34	601	7.0	811	11 Q922N2	Q922n2 mus musculu
35	599	7.0	1272	5 Q95SG4	Q95sg4 drosophila
36	585	6.8	1298	5 Q8I1E2	Q8i1e2 drosophila
37	569	6.6	1336	5 Q8I1A0	Q8i1a0 drosophila
38	542	6.3	98	11 Q8CCP4	Q8ccp4 mus musculu
39	488	5.7	673	11 Q8R0D3	Q8r0d3 mus musculu
40	488	5.7	1880	11 Q9WVPS	Q9wvps mus musculu
41	485	5.7	1021	11 Q8EUM9	Q8eum9 mus musculu
42	461	5.4	470	10 Q9FPT2	Q9fpt2 arabidopsis
43	443	5.2	465	13 Q9YH21	Q9yh21 gallus gall
44	442.5	5.2	587	13 Q9VHZ0	Q9vhz0 gallus gall
45	433	5.1	1080	11 Q9EQU1	Q9equl mus musculu

ALIGNMENTS

RESULT 1
Q8NFA0 PRELIMINARY; PRT; 1604 AA.
ID Q8NFA0
AC Q8NFA0
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Ubiquitin specific protease NYREN60.
GN NYREN60
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Paulding, A., Haber, D.,
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332301.1; J09722.1;
DR InterPro; IPR006615; DUSP.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001125; Recoverin.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00036; ehand; 2.
DR PRINTS; PR00450; RECOVERIN.
DR SMART; SM00695; DUSP; 1.
DR SMART; SM00054; EFH; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
DR PROSITE; PS00335; UCH_2_3; 1.
KW Protease.
SQ SEQUENCE 1604 AA; 181654 MW; A621F764B76321E3 CRC64;

Query Match 100.0%; Score 8563; DB 4; Length 1604;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGAKSRIGFLSYEALRRVTVLKLRLKADPKRTGSLSYNGQHCFFIREVLGDGPPKVF 60

Db 1 MGAKESRIGFISYEALRRVTDVELKELKDAFKRTCCGLSYMGCHCFIREVLGDGVPPKV 60
Qy 61 AEVYCSFGTSGKLHNNLIVGLVLLTRGKDEBKAKYIFSLFSSSGNVIIRREMERML 120
Db 61 AEVYCSFGTSGKLHNNLIVGLVLLTRGKDEBKAKYIFSLFSSSGNVIIRREMERML 120
Qy 121 HVDGKVPDTRLKCFSGEKGKVNKFNWLFNKAFAFTFSRWLLSGVYVTLTDDSDTPT 180
Db 121 HVDGKVPDTRLKCFSGEKGKVNKFNWLFNKAFAFTFSRWLLSGVYVTLTDDSDTPT 180
Qy 181 FYQTLAGVTHLESDIIDLEKRYWLLKAQRTGRFDELTEGPIVSPPIRPSISEGLFNAP 240
Db 181 FYQTLAGVTHLESDIIDLEKRYWLLKAQRTGRFDELTEGPIVSPPIRPSISEGLFNAP 240
Qy 241 DENRDNHIDPEKISCGLSACRCRGLAERQKFCFKVFDVDRDGVLSRVELRDMVVALLEVW 300
Db 241 DENRDNHIDPEKISCGLSACRCRGLAERQKFCFKVFDVDRDGVLSRVELRDMVVALLEVW 300
Qy 301 KDNRTDDIPELHMDLSIDIVEGILNAHDTTKMGHLTLEDYQIWSVKNVLANEFNLFLQVC 360
Db 301 KDNRTDDIPELHMDLSIDIVEGILNAHDTTKMGHLTLEDYQIWSVKNVLANEFNLFLQVC 360
Qy 361 HIVJGLRPATPEEGQIIRGWLERSYGLQAGHNFIIISQWQWQWKEVYKYDANPVI 420
Db 361 HIVJGLRPATPEEGQIIRGWLERSYGLQAGHNFIIISQWQWQWKEVYKYDANPVI 420
Qy 421 EPSSVLNGGKYSGFCTAAHPMEQVEDRIGSSLSYVNTTEEFSDNISTASEASTAGSGFL 480
Db 421 EPSSVLNGGKYSGFCTAAHPMEQVEDRIGSSLSYVNTTEEFSDNISTASEASTAGSGFL 480
Qy 481 YSATPGADVCFARQHTSDNNQCLLGANGNILLHNPQKPGADIDNPLVQFPVATSL 540
Db 481 YSATPGADVCFARQHTSDNNQCLLGANGNILLHNPQKPGADIDNPLVQFPVATSL 540
Qy 541 TLEGRLKRTPOLTHGRDYEMVPPVWRYALHYWGANLALPRPVINKSKTDDIPELELFR 600
Db 541 TLEGRLKRTPOLTHGRDYEMVPPVWRYALHYWGANLALPRPVINKSKTDDIPELELFR 600
Qy 601 YLLFLRQOPATRTQOSNIWNMGVPSFNAPLKRVLAYTCGFRMOTIKEIHYLSQRLR 660
Db 601 YLLFLRQOPATRTQOSNIWNMGVPSFNAPLKRVLAYTCGFRMOTIKEIHYLSQRLR 660
Qy 661 IKEEDMLWLYNSNYLTLLDDEHKLLEYLKIODEQHLVTEVRNKMSPPEMSFTANS 720
Db 661 IKEEDMLWLYNSNYLTLLDDEHKLLEYLKIODEQHLVTEVRNKMSPPEMSFTANS 720
Qy 721 KIDRHKVPTEKGATGLNLGNTCFMNSSIQCVSNTOPLTQYFISGRHLYELNRTNPIGMK 780
Db 721 KIDRHKVPTEKGATGLNLGNTCFMNSSIQCVSNTOPLTQYFISGRHLYELNRTNPIGMK 780
Qy 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTWIAKYAPFPNGFOQDQSOELLAPLLDGHED 840
Db 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTWIAKYAPFPNGFOQDQSOELLAPLLDGHED 840
Qy 841 LNRVHEKPYVELKDSGRPDWEAAEAWNDHLRNRISIVVDLPHGQLRSQVKCKTCGHIS 900
Db 841 LNRVHEKPYVELKDSGRPDWEAAEAWNDHLRNRISIVVDLPHGQLRSQVKCKTCGHIS 900
Qy 901 VRFPDFNPLSLPLMDSYMHLEITVIKLDGTTTPVRYGLRLNMDKTYGLKKQSLDLCGLN 960
Db 901 VRFPDFNPLSLPLMDSYMHLEITVIKLDGTTTPVRYGLRLNMDKTYGLKKQSLDLCGLN 960
Qy 961 SEQILLAEVHGSNKNFQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTCTDFFSSPSPT 1020
Db 961 SEQILLAEVHGSNKNFQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTCTDFFSSPSPT 1020
Qy 1021 NEMFTLTNGDLPRPIIPNGMNPVVP CGTEKXFTNGMNGMPSLPDPSPTFTYIAVH 1080
Db 1021 NEMFTLTNGDLPRPIIPNGMNPVVP CGTEKXFTNGMNGMPSLPDPSPTFTYIAVH 1080
Qy 1081 RKMWRTELYFLSSQKRNPSLFGMLIIVPCTVHTRKKDLYDAVMIQVSRSLASPLPQEASN 1140

Db 1081 RKMWRTELYFLSSQKRNPSLFGMLIIVPCTVHTRKKDLYDAVMIQVSRSLASPLPQEASN 1140
Qy 1141 HAQCDSDSMGYQYFPTLRVQKDGNSCAWCPWYFCRGCKIDCGEDRAFTGNAYIAVDWD 1200
Db 1141 HAQCDSDSMGYQYFPTLRVQKDGNSCAWCPWYFCRGCKIDCGEDRAFTGNAYIAVDWD 1200
Qy 1201 PTALHLRYQTSQERVDHEHSEVQSRRAQAEPINLDSCLRAFTSEELGENEMYCCKCK 1260
Db 1201 PTALHLRYQTSQERVDHEHSEVQSRRAQAEPINLDSCLRAFTSEELGENEMYCCKCK 1260
Qy 1261 THCLATKKLDLWRLPPLIIHLKRFQFVNGRWIKSQIKVFPRESFDPSPAPLVRPRDPAIC 1320
Db 1261 THCLATKKLDLWRLPPLIIHLKRFQFVNGRWIKSQIKVFPRESFDPSPAPLVRPRDPAIC 1320
Qy 1321 QHKPLTPGQDSELSRILAREVKVDAQSSAGEEDVLLSKSPSSLSANI1SSPKGSPSS 1380
Db 1321 QHKPLTPGQDSELSRILAREVKVDAQSSAGEEDVLLSKSPSSLSANI1SSPKGSPSS 1380
Qy 1381 RKSQTCSPSSXNSPNSPRTLGRSKGRRLRPQIGSKNKLSSSKENLDASKENGAGQICE 1440
Db 1381 RKSQTCSPSSXNSPNSPRTLGRSKGRRLRPQIGSKNKLSSSKENLDASKENGAGQICE 1440
Qy 1441 LADALSGRHVLGSSQPELVTPQDHEVALANGFLYEACNGVNGOLGNHSEEDSDTDDQ 1500
Db 1441 LADALSGRHVLGSSQPELVTPQDHEVALANGFLYEACNGVNGOLGNHSEEDSDTDDQ 1500
Qy 1501 REDTRIKPIVNLVYLAISCHSGILGGHYVYVAKNPNCKWYCYNDSSCKELHPDEIDTDSAY 1560
Db 1501 REDTRIKPIVNLVYLAISCHSGILGGHYVYVAKNPNCKWYCYNDSSCKELHPDEIDTDSAY 1560
Qy 1561 ILFVEQOQIDVIAOFLPKTDGKMADTSSMDDEDFSDYKCYVLQ 1604
Db 1561 ILFVEQOQIDVIAOFLPKTDGKMADTSSMDDEDFSDYKCYVLQ 1604

RESULT 2
Q9BX85 PRELIMINARY; PRT: 1274 AA.
ID Q9BX85
AC Q9BX85;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Ubiquitin specific protease.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sha J.H.;
RT Identification of a novel ubiquitin specific protease gene related to
RT testes development from human testes cDNA library.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
EMBL AF350251; AAK30207.1;
DR MEROPS; C19.044;
DR InterPro; IPR006615; DUSP.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR SMART; SM00695; DUSP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS50235; UCH_2_3; 1.
KW Protease.
SQ SEQUENCE 1274 AA; 143940 MW; CB84ECB7087DC58C CRC64;
Query Match 79.5%; Score 6807; DB 4; Length 1274;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 331 MGLTLEDYQIWSVKNVLANEFNLFLQVCHIVLGLRPATPEEGQIIRGWLERSYGL 390
Db 1 MGLTLEDYQIWSVKNVLANEFNLFLQVCHIVLGLRPATPEEGQIIRGWLERSYGL 60

QY 391 QAGHNFIIISMOWQWQKESVUKYDANPVVIEPSSVLNGGKYSFGTAAPMEQVEDRIGSS 450
DB 61 QAGHNFIIISMOWQWQKESVUKYDANPVVIEPSSVLNGGKYSFGTAAPMEQVEDRIGSS 120
QY 451 LSYVNTTEKFSNISTASEASETAGSGLYSATGADYCFARQHNITSDNNQCILGANG 510
DB 121 LSYVNTTEKFSNISTASEASETAGSGLYSATGADYCFARQHNITSDNNQCILGANG 180
QY 511 NILHLNPKPGAI DNOPLVTOEPVKATSLTEGGRLKBTPLI INGRDYEMPEPVWAL 570
DB 181 NILHLNPKPGAI DNOPLVTOEPVKATSLTEGGRLKBTPLI INGRDYEMPEPVWAL 240
QY 571 YHWYGANLALPRVINKSKTDIPELELFPYLLFLRQOPATRTQOSNIWVNGNVPSPNA 630
DB 241 YHWYGANLALPRVINKSKTDIPELELFPYLLFLRQOPATRTQOSNIWVNGNVPSPNA 300
QY 631 PLKRVLAYTGCSRMQTIKEIHEYL SORLRIKEEDWRLWLYNSEN YLTLLDDEHKL EYL 690
DB 301 PLKRVLAYTGCSRMQTIKEIHEYL SORLRIKEEDWRLWLYNSEN YLTLLDDEHKL EYL 360
QY 691 KIODEOHLVTEVRNKMDSWPEEMSFANSKIDRHKVPTEKGATGLSNLGNTCFVNSSIQ 750
DB 361 KIODEOHLVTEVRNKMDSWPEEMSFANSKIDRHKVPTEKGATGLSNLGNTCFVNSSIQ 420
QY 751 CVSNTOPLQYFTSGRHLVELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRT 810
DB 421 CVSNTOPLQYFTSGRHLVELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRT 480
QY 811 IAKYAPRNGFOODSOELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWND 870
DB 481 IAKYAPRNGFOODSOELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWND 540
QY 871 HLNRNSIVVDLPHGQLRSQVKCTCGHISVRPDPNF.LSLPLPMDSYMHLEITVVKLDG 930
DB 541 HLNRNSIVVDLPHGQLRSQVKCTCGHISVRPDPNF.LSLPLPMDSYMHLEITVVKLDG 600
QY 931 TTPVRYGLRLNMDKTYGLKQISDLCGLNSEOILLAEVHGSNIKNFPQDNQKVLRSVSG 990
DB 601 TTPVRYGLRLNMDKTYGLKQISDLCGLNSEOILLAEVHGSNIKNFPQDNQKVLRSVSG 660
QY 991 FLCAFAIPVPSISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCG 1050
DB 661 FLCAFAIPVPSISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCG 720
QY 1051 TEKNFTNGMNGHMPSLPDSPTGYIIA VRKQMRTELIFLSSQKRP.SLFGMPLI.VPCT 1110
DB 721 TEKNFTNGMNGHMPSLPDSPTGYIIA VRKQMRTELIFLSSQKRP.SLFGMPLI.VPCT 780
QY 1111 VHTRKDLYDAVMOVSRLASPLPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWC 1170
DB 781 VHTRKDLYDAVMOVSRLASPLPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWC 840
QY 1171 PWYRFGCKIDCEDRAFIGNAYIAVMDPTALHRYQTSQBRVVDEHESVEQSRAQA 1230
DB 841 PWYRFGCKIDCEDRAFIGNAYIAVMDPTALHRYQTSQBRVVDEHESVEQSRAQA 900
QY 1231 EPINLDSCLRAFTSEELGENEMYCSKCTHCLATKLDLWELP.TLIIHLKRFQVNG 1290
DB 901 EPINLDSCLRAFTSEELGENEMYCSKCTHCLATKLDLWELP.TLIIHLKRFQVNG 960
QY 1291 RWIKSQXIVKFPRESFDPSPALCOHKLPTPQGDSELPRIAREVKVDAQSS 1350
DB 961 RWIKSQXIVKFPRESFDPSPALCOHKLPTPQGDSELPRIAREVKVDAQSS 1020
QY 1351 AGEDVLLSKPSLSANI.ISSPKGSPSSRKSGTSCPSSKNSSPNSSPRTLGRSGRLR 1410
DB 1021 AGEDVLLSKPSLSANI.ISSPKGSPSSRKSGTSCPSSKNSSPNSSPRTLGRSGRLR 1080
QY 1411 LPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVGLGSGQSELVTPQDHEVALAN 1470
DB 1081 LPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVGLGSGQSELVTPQDHEVALAN 1140

QY 1471 GFLYEACNGYSGOLGNHSEEDSTDDORETRIKPIYNLYAISCHGILGGHYVY 1530
DB 1141 GFLYEACNGYSGOLGNHSEEDSTDDORETRIKPIYNLYAISCHGILGGHYVY 1200
QY 1531 AKNPCKWYCYNDSSCKELHPDIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSMD 1590
DB 1201 AKNPCKWYCYNDSSCKELHPDIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSMD 1260
QY 1591 EDPESDYKKYCVLQ 1604
DB 1261 EDPESDYKKYCVLQ 1274
RESULT 3
Q81WT4
ID Q81WT4 PRELIMINARY; PRT; 1406 AA.
AC Q81WT4
DP 01-MAR-2003 (TREMUREL. 23, Created)
DT 01-MAR-2003 (TREMUREL. 23, last sequence update)
DT 01-MAR-2003 (TREMUREL. 23, last annotation update)
DE Ubiquitin-specific protease USP6.
GN USP6
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Paulding C., Haber D.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY143550; AAN38838.1; -
KW Protease.
SQ SEQUENCE 1406 AA; 158657 MW; D3A6822CEB441DB3 CRC64;
Query Match 53.1%; Score 4544; DB 4; Length 1406;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 853; Conservative 21; Mismatches 30; Indels 4; Gaps 1;
QY 701 EVRKNDMSWPEEMSFANSKIDRHKVPTEKGATGLSNLGNTCFVNSSIQCVSNTOPLTQ 760
DB 499 EVHNKMSWPEEMSFANSKIDRKVPTEKGATGLSNLGNTCFVNSSIQCVSNTOPLTQ 558
QY 761 YFISGRHLYELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRTIAKYAPRFG 820
DB 559 YFISGRHLYELNRTNPIGMKGHMAKCYGDLVQELWSGTQKVAPLKLRTIAKYAPRFG 618
QY 821 FQOQDSOELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWNDHLNRNSIIV 880
DB 619 FQOQDSOELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWNDHLNRNSIIV 678
QY 881 DLPHGQLRSQVKCTCGHISVRPDPNF.LSLPLPMDSYMHLEITVVKLDGTPVRYGLRL 940
DB 679 DLPHGQLRSQVKCTCGHISVRPDPNF.LSLPLPMDSYMHLEITVVKLDGTPVRYGLRL 738
QY 941 NMDEKYTGLKQISDLCGLNSEOILLAEVHGSNIKNFPQDNQKVLRSVSGFLCAFEIPVP 1000
DB 739 NMDEKYTGLKQISDLCGLNSEOILLAEVHGSNIKNFPQDNQKVLRSVSGFLCAFEIPVP 798
QY 1001 VSPISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCGTEKNFTNGMV 1060
DB 799 VSPISASSPTQTDSSSPSTNEMFTLTNGDLPRIPIFNGMNTVVPVCGTEKNFTNGMV 858
QY 1061 NGHMPSLPDSPTGYIIA VRKQMRTELIFLSSQKRP.SLFGMPLI.VPCTVHTRKDLYD 1120
DB 859 NGHMPSLPDSPTGYIIA VRKQMRTELIFLSSQKRP.SLFGMPLI.VPCTVHTRKDLYD 918
QY 1121 AVWQVSLASPLPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWCWYRFGCK 1180
DB 919 AVWQVSLASPLPQASNAHQD.CDDSMGYQYPTFLRVVQKDGNSCWCWYRFGCK 978
QY 1181 IDCGRDRAFI GNAYIAVMDPTALHRYQTSQBRVVDEHESVEQSRAQAEPINLDSCLR 1240
DB 979 IDCGRDRAFI GNAYIAVMDPTALHRYQTSQBRVVDEHESVEQSRAQAEPINLDSCLR 1038

```
QY 1241 APTSEELGENEYKCKKTHCLATKLDLWFLPPIIIHLKRFQVNGRWIKSQKIVK 1300
DB 1039 APTSEELGESEMYKCKKTHCLATKLDLWFLPPIIIHLKRFQVNDQWIKSQKIVR 1098
QY 1301 FPRESDFSAFLVPRPALCOHKPLTPQGEUSEPRILAREVKVVDQAQSSAGEEDVLSK 1360
DB 1099 FLRESDFSAFLVPRPALCOHKPLTPQGEUSEPRILAREVKVVDQAQSSAGEEDVLSK 1158
QY 1361 SPSSLANIISPKGSPSSSRKSGTSCPSKSSPNSSPRTIGRSKGRRLRLOIGSKNKL 1420
DB 1159 SPSSLANIISPKGSPSSSRKSGTSCPSKSSPNSSPRTIGRSKGRRLRLOIGSKNKP 1218
QY 1421 SSSKENLDASKENGAGQICELADALRGHVLGSGSQPELVTPQDHEVALANGFLYEACG 1480
DB 1219 SSSKENLDASKENGAGQICELADALRGHVLGSGSQPELVTPQDHEVALANGFLYEACG 1278
QY 1481 N----GYSNGOLGNHSEEDSTDQREDTHIKPIYNIYALSCHSGILSGGHYITYAKNPNC 1536
DB 1279 NGCGDYSNGOLGNHSEEDSTDQREDTHIKPIYNIYALSCHSGILSGGHYITYAKNPNC 1338
QY 1537 KWCYNDSSCKELHPDEIDTDSAYILFYEQQGDIDYAFKPKTDGKQWADTSSMDDEDFFSD 1596
DB 1339 KWCYNDSSCKELHPDEIDTDSAYILFYEQQGDIDYAFKPKTDGKQWADTSSMDDEDFFSD 1398
QY 1597 YKXVCVLQ 1604
DB 1399 YEKYSMLQ 1406

RESULT 4
Q9Y591 PRELIMINARY; PRT; 828 AA.
AC Q9Y591;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NY-REN-60 antigen (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX "Antigens recognized by autologous antibody in patients with renal-
cell carcinoma."
RL Int. J. Cancer 83:456-464(1999).
DR 'EMBL: AF155116; AAD42882.1; -.
DR MEROPS; C19.044; -.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS50235; UCH_2_3; 1.
FT NON TER 1
FT NON TER 828
SQ SEQUENCE 828 AA; 94806 MW; EEB03264D50DDCC2 CRC64;

Query Match 51.9%; Score 4440; DB 4; Length 828;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 536 KATSLTLEGGRKXKTPOLIGHRDYEMVPEPVWALVHWYCANALPRPVKSKTDIPEL 595
DB 1 KATSLTLEGGRKXKTPOLIGHRDYEMVPEPVWALVHWYCANALPRPVKSKTDIPEL 60
QY 596 ELFPYLLFLRQQPATRTOQSNINWNGVSPNAPLKRVLAYTGCFSRMQTIKEIHEYL 655
DB 61 ELFPYLLFLRQQPATRTOQSNINWNGVSPNAPLKRVLAYTGCFSRMQTIKEIHEYL 120
```

```
QY 656 SQRIRKEEDRWLWLNSENYLTLLDDEHKLBYLKIQDEQHLVIEVRNKMDSWPBEMSF 715
DB 121 SQRIRKEEDRWLWLNSENYLTLLDDEHKLBYLKIQDEQHLVIEVRNKMDSWPBEMSF 180
QY 716 IANSKIDRHKVPTKCATGSLNLTNTCFNNSIQCVSNTQPLTOYFISGRHLYELNRTN 775
DB 181 IANSKIDRHKVPTKCATGSLNLTNTCFNNSIQCVSNTQPLTOYFISGRHLYELNRTN 240
QY 776 PIGMKHMAKCYGLDVLQELWSGTQKNVAPLKLWTIAKYAPRNGFQOQDSQELLAPLLD 835
DB 241 PIGMKHMAKCYGLDVLQELWSGTQKNVAPLKLWTIAKYAPRNGFQOQDSQELLAPLLD 300
QY 836 GLHEDLRVHEKPYVELKQSDGRPDWEVAAEANDNLRRNRSIVDLFHQQLRSQVKCKT 895
DB 301 GLHEDLRVHEKPYVELKQSDGRPDWEVAAEANDNLRRNRSIVDLFHQQLRSQVKCKT 360
QY 896 CGHLSVRFDFPNFLSLPLPMDSYMHLEITVILKDGTTFVRYGLRLNMDKVTGLKKQLSD 955
DB 361 CGHLSVRFDFPNFLSLPLPMDSYMHLEITVILKDGTTFVRYGLRLNMDKVTGLKKQLSD 420
QY 956 LCGLNSQIILAEVHGSIKNFPQDNQKVLSSVSGFLCAFEIPVPVSPISASSPTQDFFS 1015
DB 421 LCGLNSQIILAEVHGSIKNFPQDNQKVLSSVSGFLCAFEIPVPVSPISASSPTQDFFS 480
QY 1016 SSPSTNEMFTLTNGDLPRDIFIPNGMPNTVPCGTEKNTGKMGHMBESLPDSPTGY 1075
DB 481 SSPSTNEMFTLTNGDLPRDIFIPNGMPNTVPCGTEKNTGKMGHMBESLPDSPTGY 540
QY 1076 IIAVHRKQMTELYFLSSQKQNPFLFCMPLIVPCTVHTRKKLDYDAWIOVSRLASPLPP 1135
DB 541 IIAVHRKQMTELYFLSSQKQNPFLFCMPLIVPCTVHTRKKLDYDAWIOVSRLASPLPP 600
QY 1136 QEASNAHQDQDSDSGYQYPTFLRVQKDGNSCAWCPWYRCRCCKIDCGEDRAFIGNAYI 1195
DB 601 QEASNAHQDQDSDSGYQYPTFLRVQKDGNSCAWCPWYRCRCCKIDCGEDRAFIGNAYI 660
QY 1196 AVDWDPALHLRYOTSQERVVDEHESVEQSRRAQAEIPINLDSCLRAFTSEELGENEMY 1255
DB 561 AVDWDPALHLRYOTSQERVVDEHESVEQSRRAQAEIPINLDSCLRAFTSEELGENEMY 720
QY 1256 CSKCKTHCLATKLDLWFLPPIIIHLKRFQVNGRWIKSQKIVKPFRESDFSAFLVPR 1315
DB 721 CSKCKTHCLATKLDLWFLPPIIIHLKRFQVNGRWIKSQKIVKPFRESDFSAFLVPR 780
QY 1316 DPALCQHKPLTPQGEUSEPRILAREVKVVDQAQSSAGEEDVLSKSPS 1363
DB 781 DPALCQHKPLTPQGEUSEPRILAREVKVVDQAQSSAGEEDVLSKSPS 828

RESULT 5
Q9VM49 PRELIMINARY; PRT; 1736 AA.
AC Q9VM49;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG8334 protein.
GN CG8334.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1] SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```


OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RA Goubrin F., Dominguez J., Draetta G.F.;
RT "cDNA cloning of an isoform of the human tre oncogene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ012755; CAA10171.1; --
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
SQ SEQUENCE 354 AA; 39586 MW; 7112708E118EC127 CRC64;

Query Match 19.5%; Score 1668; DB 4; Length 354;
Best Local Similarity 92.7%; Pred. No. 4 4e-117;
Matches 316; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

QY 1264 LATKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVKFPRESFDPFSAFLVPRDPALCOHK 1323
DB 14 LATKLDLWRLPPFLIIHLKRFQVNDQWIKSQKIVKFPRESFDPFSAFLVPRDPALCOHQ 73
QY 1324 PLTPQDELSEPRILAREVKYKVDQSSAGEEDVLISKSPSSLSANTISSPKGSPSSSRKS 1383
DB 74 PLTPQDELSEPRILAREVKYKVDQSSAGEEDVLISKSPSSLSANTISSPKGSPSSSRKS 133
QY 1384 GTSCPSKSNSSPNSSPRTILGRSGRLRFPQIGSKNKLSSKENLDASKENGAGQICELAD 1443
DB 134 GTSCPSKSNSSPNSSPRTILGRSGRLRFPQIGSKNKLSSKENLDASKENGAGQICELAD 193
QY 1444 ALSRGHVGSGQPELVTPQDHEVALANGFLYEACGNGYNGOLGNHSEEDSTDQDRED 1503
DB 194 ALSRGHVGSGQPELVTPQDHEVALANGFLYEACGNGYNGOLGNHSEEDSTDQDRED 253
QY 1504 TRIKPIYNLYAISCHSGILGGHYVYAKNPCKWYCYNDSSCKELHPDEITDTSAYILF 1563
DB 254 THIKPIYNLYAISCHSGILGGHYVYAKNPCKWYCYNDSSCKELHPDEITDTSAYILF 313
QY 1564 YEQQIDYAOFLPKDTGKQKMDTSSMDEDFESDYKYCVLQ 1604
DB 314 YEQQIDYAOFLPKDTGKQKMDTSSMDEDFESDYKYCVLQ 354

RESULT 8

Q15634 Q15634 PRELIMINARY; PRT; 786 AA.
AC Q15634;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Oncogene.
GN TRE.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ewing's sarcoma;
RX MEDLINE=92228503; PubMed=1565468;
RA Nakamura T., Hillova J., Mariage-Samson R., Onno M., Huebner K.,
RA Cannizzaro L.A., Boghosian-Sell L., Croce C.M., Hill M.;
RT "A novel transcriptional unit of the tre oncogene widely expressed in
human cancer cells";
RL Oncogene 7:733-741(1992).
DR EMBL; X63546; CAA45108.1; --
DR InterPro; IPR000195; RabGAP_TBC.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00566; TBC; 1.
DR Pfam; PF00443; UCH; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS0086; TBC_RABGAP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS0235; UCH_2_3; 1.
SQ SEQUENCE 786 AA; 83546 MW; CC0C9280ED15DA0A CRC64;

Query Match 16.4%; Score 1406; DB 4; Length 786;
Best Local Similarity 92.7%; Pred. No. 9 4e-97;
Matches 266; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 701 EVRNKMSWPEEMSFANSKIDRHKVPTKGTGLSNLNTCFMSSIOCVSNTPQLTQ 760
DB 499 EVRNKMSWPEEMSFANSKIDRHKVPTKGTGLSNLNTCFMSSIOCVSNTPQLTQ 558
QY 761 YFISGRHLYELNRTNPTGMKGMAKCYGDIQVBLWSGTQKNVAPLKLRTIAYAPFNG 820
DB 559 YFISGRHLYELNRTNPTGMKGMAKCYGDIQVBLWSGTQKNVAPLKLRTIAYAPFNG 618
QY 821 FQODSOELLAFLDGLGHEDLNVRHEKPYVELKDSGRPDWEVAAEAWDNHLSRSIVV 880
DB 619 FQODSOELLAFLDGLGHEDLNVRHEKPYVELKDSGRPDWEVAAEAWDNHLSRSIVV 678
QY 881 DLHFQQLRSQVKCTCGHISVRFPDFFNLSPLPMDSYHLEITVTKDGTTPVRYGLRL 940
DB 679 DLHFQQLRSQVKCTCGHISVRFPDFFNLSPLPMDSYHLEITVTKDGTTPVRYGLRL 738
QY 941 NMDEKYTLGKQLSDICGLNSEQILLAEVHGSNIKPPQDNQKVRLS 987
DB 739 NMDEKYTLGKQLSDICGLNSEQILLAEVHGSNIKPPQDNQKVRLS 785

RESULT 9

Q8WT43 Q8WT43 PRELIMINARY; PRT; 1175 AA.
AC Q8WT43;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 133.8 kDa protein (Ubiquitin c-terminal hydrolase).
GN ZK328.1 OR CYK-3.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favello A.;
RT "The sequence of C. elegans cosmid ZK328.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]

RP SEQUENCE FROM N.A.
RA Kaitna S., Schnabel H., Schnabel R., Hyman A.A., Glotzer M.;
RT "An ubiquitin C-terminal hydrolase is required to maintain osmotic
balance and execute actin-dependent processes in the early
Caenorhabditis elegans embryo.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50193; AAL32262.1; --
DR EMBL; AF469173; AAL79016.1; --
DR WormPep; ZK328.1a; CE30061.
DR InterPro; IPR008615; DUSP.
DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR001394; UCH-2.

DR Pfam; PF00036; UCH-2.

DR Pfam; PF00443; UCH-1.

DR SMART; SM00695; DUSP; 1.

DR SMART; SM00054; EFH; 2.

DR PROSITE; PS00018; EF HAND; 2.

DR PROSITE; PS0235; UCH 2_3; 1.

KW Hypothetical protein; Hydrolase.

SQ SEQUENCE 1175 AA; 133832 MW; 734DE99C5ABF77A7 CRC64;

Query Match 14.4%; Score 1233; DB 5; Length 1175;

Best Local Similarity 24.2%; Pred. No. 2e-83;

Matches 392; Conservative 216; Mismatches 487; Indels 528; Gaps 43;

QY 11 LSVVEALRVTVDELKRLKDAFKRCCGLSYNGQCFIREVLGDC-VPPKVAEVIYCSFG 69

DB 16 ISAEADAKTVISDEEYRIRQAQRFQNGCINDEFY--HVLGGAQIPEEKRLIFSFFS 73

QY 70 GTSKGLHFNLLVGLVLLTRGKDEKAKYIFSLFSESGNYVIREEMERMLHVVQKVPD 129

DB 74 HGAETISFDNLLSSLVGLCR-VEEVQSRPI-----EE----- 104

QY 130 TLKCFSEGEKYNKFRNWLFLNKDAFTFSRWLLSGV--YVTLTDDSDTPFYQTLA 186

DB 105 -----YHEFASW-----GLSPPKLTIPLNDYISIFYEVM 134

QY 187 GYTHLEESDIIDLEKRYMLLKQASRTGRPDLETFGPLVSPPIRPSLSGLFNADENRDN 246

DB 135 YVTHLSVNEVIELEKVFATISDRA-VCKLNEEKWQALGCGPPDSYARLFAVFDENRDG 193

QY 247 HIDEKISGLGACGRCGLAERQKCFKVDVDRGVLSRVELRDMVVALLEVKNRPTD 306

DB 194 QIDFRELVTLSALCRGPFGRISOLARIWVDDCKLSDSELSNM-----YKDL--- 243

QY 307 DIPELHMDLSDIVEGILNAHDTTKMGHTLTDYQIWSVKN-VLANEFLNLLFQVCHIVLG 365

DB 244 NVPEEHQTVT-----KSSNGKSALVDFGIWAQENKYNVEYYSMAIQGHICLG 292

QY 366 LRPAPEEGQIIRWLEERSYGLQAGHNWIIISQWQQWKEVYKYDANPVIEPSSV 425

DB 293 LRPEGRKMLQIWNFEKRASELPLS---EWNIVASGWHAEIERSFLEADKNPNPINDSGI 349

QY 426 LGGKYSFGTAAPMEQVEDRIGSSLSYVNTVEEFSDNISITASEASETAGSGFLYSATP 485

DB 350 -----KGTREDGWS----- 359

QY 486 GADVCFARHNTSDNNQCLLGANGNILLNLPQKGAIDNQLVTPQPVKATSLTEGG 545

DB 360 -----KVACISABSA 369

QY 546 RLKRTPOLIHGRDYEMVPPVWRALYHWYGANLALPRPVKN-----SKTDIPELELFP 599

DB 370 RLK--PDLLPS-DYIRVPVPLWRMLRWGCGALTVDQSQTRKYLDGEFEEDKNPALELYP 426

QY 600 RYLLFL---RQOPATRTQOSNIWNGNVPSPNAPLKVLYATGC--FSRMQTIKEIHEY 654

DB 427 LEILLGLHGRKKSQDGT-----NTP--RSLTSWACAQVSRSMVTDELLAL 470

QY 655 LSQRLAIKEEDMELWLYNSNY--LTLDDDEHKLLEYL-----KIODEQHLVLEVRNK 705

DB 471 CKTELGLGDARLQVWVNEEGNVLDDGQNLHLQYSSLGKTKYKNKMLLEVER 530

QY 706 DMS-WPEEM-----SFIANSKIDRHKVKGATGLSNLGNCTCFWNSSTIQCVSNTQP 757

DB 531 GTGVWPEELRASLSGQITAASTLSNAQLSGRPGAVLVNNGNFCYRNGAQICLARVSP 590

QY 758 LTOYFISGRHLYELNRTNPIGMKGMKAC---YGLVQELNSGTQKNVAPLKLRTIAYK 814

DB 591 LTOYFIDEDNDIAKGNL--RRGDAETITIEYAKLLREMWAAKKKNIAPNDFNDAILKS 648

QY 815 APRFNGFQQQDSQELLAFLLDGLHEDLNVRHKKPYVELXDSG----- 857

DB 649 SDMFECSEQHDCQCFVAFLLDQLTSMYBSNKSLSHPSESEGTDSNKLSDSKKKEADK 708

QY 858 -RPMWEVAAEAMDNHLRRNRSIVVDLHFHQLSRQVKCKTCGHISVRFDFPNFLSLPLMD 916

DB 709 EADEEKAERSWTEYKQNESLVTQLTQLRLICRTQSSSSVFEPFELSPLDIGE 768

QY 917 SYMHLEITVTKLDGTTVPYRGLRNLNDEKTYGLKKQLSDLCGLNSEQIILAEVHGSNKN 976

DB 769 DVDLYQVIVVHRDGRIPRAYGFRLSRDSKVGNLREVAVSSGIGSMHLTIQ----- 819

QY 977 FPQDNQKRLSVSGFLCAFEIPVPVPSIASASPTQTD--FSSSPSTNEMFTLTNGDLPR 1034

DB 820 -----CMSSKGTLL-----MSRSPNRRSNLRDELPLSPFPFGARLYAL----- 857

QY 1035 PIPNGMNTVVPCCTEKNFTNGMVGHPMLSPSPFTG---YIIAIVRRKQM-RTELYP 1090

DB 858 -----ELPES--TGEDQWRVAMRKLQYNHEPYI 884

QY 1091 LSSQKN-RPSLFQMPILIVPCTVHTKRLKOLYDAWTVQSVRLASPLPQEBASNAHQDCDDSM 1149

DB 885 LGSTAGFIVSRFGLPLIVGLDEEVTGKLYEDVMYQMRHFMHESVNSSSRADHPCEDE- 943

QY 1150 GYCPPTLRVQKDGNSCAWCPWYRRCRCCKIDCGEDRAFI-GNAYIAVDWDPHTALHLY 1208

DB 944 NSGYPTLCLVDENYWCQCPCALRRCRCPIRPPDESKVFI PANCEFLAVDWLPALYLY 1003

QY 1209 QTSQERVDEHESVQSRRAQAEPIINDSCLRAFTSEELGENEMYCKKTHCLATKK 1268

DB 1004 NBSQEQACEDDPSVAETWSRHFPAPSLSEHCIEKFCFETL--DAATQCDCRCKKTRDKV 1061

QY 1269 LDWLRLPPLIHLKRFQV--NGRWIKSQKIVKPPRESFDSAPLVRPDPALCQHKPLT 1326

DB 1062 MTWKLPKYLIIHLKPEFLRQGRMGKCKRTVNFPLKHFDPAPF----- 1106

QY 1327 PQGDELSEPRILAREVKYVDAOSSAGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTS 1386

DB 1107 -----VDRPDGN----- 1113

QY 1387 CPSSKNSSPNSSPRTLGRSKGRLLRPLQIGSKNKLSSSKENLDASKENGAGQICELADALS 1446

DB 1114 ----- 1113

QY 1447 RGHVLSQSPELVTPQDHEVALANGFLYEACNGYNGQLGNHSEEDSTDQREDTRI 1506

DB 1114 ----- 1113

QY 1507 KPIYNIYAISCHSGIILGGHYVITYAKPNCKWYCNVDSCKELHPDEIDTSAIILFYEQ 1566

DB 1114 --TYECIALANHYGQLSCGHFTIAYAKSNEDKWLILLNDCSVREVSEBVDKQGYLLFYER 1171

QY 1567 QGI 1569

DB 1172 KDV 1174

RESULT 10

Q8WT44

ID Q8WT44 PRELIMINARY; PRT; 1178 AA.

AC Q8WT44;

DT 01-MAR-2002 (T-EMBLrel. 20, Created)

DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Hypothetical 134.1 kDa protein.

GN ZK328.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for


```

DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS02035; UCH_2_3; 1.
KW Nuclear protein.
SQ SEQUENCE 915 AA; 103752 MW; 1FF0537D187D4904 CRC64;

      Query Match      12.8%; Score 1095.5; DB 11; Length 915;
      Best Local Similarity 24.8%; Pred. No. 3.1e-73;
      Matches 318; Conservative 166; Mismatches 323; Indels 477; Gaps 37;

Qy 390 LQAHNWFILSMQWQWQKVKYKDYANVPVIEBPSVLVNGGKYGISFGTAAPMEQVEDRIGS 449
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28 LQGAQVYILDSRWFQKWKYVGF-----SWDYNVG----- 60

Qy 450 SLSVNTTEKFSNIDTASEASETAGSGFLYSATPGADVCFARQHNSTDNNNQCLLGAN 509
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 -----EHLF----- 65

Qy 510 GNILLHLNPKPGAIDNPQIVTQBPVKATSLTLEGGRLKRTPOLIHGRDVMYVPEPVRA 569
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 -----PGPIDNSGLFS-DP-----ESQTLKE--HLIDELDVLVPAEAWNK 103

Qy 570 LYHWYGANLALPRVIKNSKTDIPELELPR-----YLLFLRQOPATPRQQSNIVNWNG 623
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
104 LLNWYGC-VEGQQITVRK-----VVEHGLFWKCKVEYVLELKK----- 142

Qy 624 NVPSNPAPLKRVLAYTGC-FSRMOTIKEIHYLSQRLRI-KEEDMRLLWYNSENYLTLLD 681
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 ---CENSDDPTNVLIS---CHFSKADTIATIEKMRKLFNIPAEERETRLWNKYMSNTVEQLS 196

Qy 682 DEDHKLVELKIQDQHLVIVERNKDMWPEMSPIAN-----SSKIDRHKYPT---EKGAT 734
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
197 KLDNTIQAGLYQQQVLVIRPQNEDEGWPKQ-SLQSNQSGFSASYNQCEPSPSHIQGLC 255

Qy 735 GLSLNGNTCFMNSSIQCVSNTQPLTYFISGRHLYELNRTNPIGMKHMACKYCGDLVLEL 794
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
256 GLGLNGNTCFMNSALQCLSNAPTLEYFLKDEYEAENRDNPLCMKGEIAEYAEILKQM 315

Qy 795 WSGTQKNVAPLKRWTAKYAPRNGSQODSQQELLAFILDLGLHEDLNRVHEKPYVELKD 854
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
316 WSGRDTTHVAPMFTQVGRPAQPSGQQODSQELLAFILDLGLHEDLNRVKKPKYLEPKD 375

Qy 855 SDGRPDVEAAEANDHLNRNRSTVVDLFHQQLRSQVKCKTCGHISVRFDPFPNPLSLPLP 914
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
376 ANGRPDVAVAKAENHRLNDSVIVDTFHGLFKSLTIVCPKCAKVSVTDFPFCVLTPLP 435

Qy 915 MDSYWHLEITVILKD-GTTPVRYGLRLNMBEKYTLGKQLSDLCGLNSEQILLAEVHGSN 973
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
436 LKQDRIMEVFLVPADPCRPQIYETVPLMGAISDLCEALSKLSGIAAENNVVTDVYNHR 495

Qy 974 I-KNFPQDNQKVLRSVSGFLCAFEIPVPSPIASSTPTQDFFSSPTNEMFTLTNGDL 1032
Db   496 FHKIFQMDGLSHITPRDDIFVYEV-----CNMSMDGSECITL----- 533

Qy 1033 PRPIFINGMNTVVPCTEKNFTNGVMNGHMPSLDPSPTGYTIAVHRKMRTELTYFLS 1092
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
534 --PYV-----REKK-----SRPSSASSGAV----- 552

Qy 1093 SOKNRPSLFGMLPIVPCTVH-TRKLDLYDAYWIOVSR-----LAGPLP----- 1135
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
553 -----LYGQLLVSVPKKLTLESLYQAVCDRIKSYIKQPLPDEFLSPLEPCACNGS 605

Qy 1136 -----QASNAHQD-----CDDSMG-----VOYPTFLRVQ 1161
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
606 RSSYEGDEEEMDHQEGKEQLSEVEGSGENDQGDHSHESAQKVKGQOPRHKRLFTFSLV- 664

Qy 1162 KDGNSCAWCPWYRCRGCKIDCGEDRAFIGNAYIAVDMDPTALHLRYOTSERVVDSHES 1221
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
665 ---NSCGTADINSLATDGL-----LKNSRSTIAIDWDSSTRSLYDFDQSEACEKHL 716

Qy 1222 VEQSRRAQAEPIINLDSCLRAFTSBEELGENENMYCSKCTHCLATKCLDLWRLPILIIH 1281
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
717 MSQPOKQKAAVAIRCEIETFTMETLGEHPWTCPTCKHQOATKFPDLMSLPSKILAVH 776

```

663 BEDMLWLNSENYLTLLDDEHKLLEYLKIODEQHVLVIEVRNKMDSWPBEEMS----- 714
 198 EYETRLUNRWNTYEQLTKESTVODAGLYQGMALIEKRNADGTWPRQSTKSTATS 257
 715 -----FIANSKIDR-----HKVPTK 731
 258 RNPATSGKPTSSYSTSSSSMTNGSDSTHALNNSGLNKGGYTPSYSSYRESSSQP 317
 732 GATGLNLTGNTCPMSSICVNTQPTQYFISGRHLYELNRTNPIGMKGMKACVGLV 791
 318 GUGLNLGNTCPMSSALCLNSPLTRYFUTDEYKKEINRENPLGMKEIAYAYALI 377
 792 QELWSGTQKNVAPLKLRTIAKVPAPNGFQQQDSQELLAFLLDGLHEDLNVRHVKPYVE 851
 378 RQIWSGDHSVAPRPFKTVQGRFAPQFSGVQQQDSQELLAFLLDGLHEDLNVRKPKYLE 437
 852 LKSDGRPDWEVAAEAWDHLNRNRSIVVDLPHGQLRSQVCKTCCHISVRDPNFI 911
 438 LKANGRPDSIVAAEAWENHLNRNRSIIVDIHGLFKSTLVCPCHKVSVTFDPFCYLT 497
 912 PLPMDSYMHEITVIKLD-GTTPVRYGLRLNMDKTYGLKQLSDCLGNSQIILAEVH 970
 498 PLPMKDKORTWEVFRVSDPHCKPQVRLVPRKGVVLDLCSLSKISDVQAEVWVTVY 557
 971 GSN-KNFPQDNQKRLSVSGFLCAFEIIPVPSIASPTQTDSSSPSTNMTFTLTN 1029
 558 NHRFHKIFQDEMLSHIMBRDDIFVYE---VKPRXA-----EDWVTL--- 596
 1030 GDLPRPIFIPNGMPTVVPQGTETKNTNGMNGHMPSLPDSPTGTGYIIAVERKQWRTELY 1089
 597 ---PI---CFRERRRQATSTGV----- 614
 1090 FLSSQKNRPSLFQMPILVPCVTHTRKD-LYDAVWIOVSR-LASPLP-----POEAS 1139
 615 -----LFQOPLMISVPKHNLTLDILXENVLDRIKRYVKNPPELYTCLSDSEAC 664
 1140 NHAQ-DCDDSMGY-----QYFPTLRVVQKDNWSCAWCPWRFCEGK--- 1180
 665 NGSNGCEDDMDHOEBQESQETNGQDSVEEENSEGG-----CSPKNMPK 714
 1181 --IDCEDRAF--IGNAY-----IAVMDPTALHLYRTQSER 1214
 715 KALEC-RKRLFTSLVNSYATAEVSMPAKGYLKLSFATVALDWSBSKKLFYDDQBAE 773
 1215 VDEHRSVEQSRQAEPINLDCSLRAFTSEBELGENEMYCKKTHCLATKGLDLRL 1274
 774 AFDKHESSMMQPKKV-TVALRECIOFTKTTETLGEHDPWPCNCKKHQATKFDLWSL 832
 1275 PPIILHLKRFQFVNGRW-IKSQKIVKFPRESFDPASFLVPRDPALCQHKPLTPQDELS 1333
 833 PKILVHLKFPY-NRFRWKLDAVVEFPIRLNMSEF----- 869
 1334 EPRILAREVKVDAQSAGEEDVLLSKSPSSLANTISSPKGSPSSSRKSGTSCPSSKNS 1393
 870 -----VCDPQAGFF----- 878
 1394 SPNSSPTLGRSKRLPLPOIGSKNKLSSKENLASKENGAGQICELADALSRGHVLOG 1453
 879 ----- 878
 1454 SQPELVTPQDHEVALANGFLYHEACNGYSNGQLGNHSEEDSTDQREDTRIKPIYNLY 1513
 879 -----VYDLV 883
 1514 AISCHSILGGHVVYAKNPN-CKWYCYNDSSCKELHPPDEIDTDSAYILFEEQ-GIDY 1571
 884 AVSDHYGMMGGHYTAYAKQDKQWYFDDSSPASSEDQIVTKAAYVLYFORQDNVDC 943
 1572 AQFLPKTDGKQWATSSMDSEDFE 1594
 944 KNSSPAAS-----IGTPAVEEDLE 962

RESULT 14
 Q8BTL9
 ID Q8BTL9 PRELIMINARY; PRT; 962 AA.
 AC Q8BTL9;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Ubiquitin specific protease 4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; PubMed=12466851;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK089425; BAC40877.1; -;
 SQ SEQUENCE 962 AA; 108342 MW; 95F95BB86186DA52 CRC64;
 Query Match 12.7%; Score 1086; DB.11; Length 962;
 Best Local Similarity 24.2%; Pred. No. 1.7e-72;
 Matches 322; Conservative 165; Mismatches 321; Indels 522; Gaps 38;
 390 LQAGNWFIIISMOWQOWKEYVKYDANPVVIEPSSVLNGKGYSGFTAHPMEQVEDRIGS 449
 28 LQGAQWYLLDSRWFKQWKKYGVGD-----SWMNVVG----- 60
 450 SLSVYNTTEKFSFNISTASEASETAGSGFLYATPGADVCFAROHNSTNNNOCLLGAN 509
 61 -----EHLNF----- 65
 510 GNILLHLNPKGAIQDNOPLVTOEPVKATSLTLEGGELKRTPOLIHRDYMVPVWRA 569
 66 -----PGIDNSGLFS-DP-----ESQTLKE-HLIDELDIVLPAEAWNK 103
 570 LYHWYGANLALPRPVIKNSKTDIPELELPPR-----YLLFLQOQATRTQQSNINWVMG 623
 104 LLNMYGC-VEGQQITVRK---VVEHGLFVKHCKVEYVLELKL----- 142
 624 NVPSPNAPLKRVLAYTCG-FSRMQTIKEIHEYLSQRLRI-KEEDMLRLWLNSENYLTLD 681
 143 ---CENSDDPTNVLVS---CHFSAKDTIATIEKMKULFNIIPAERETRLNWKYNTYEQLS 196
 682 DEDHKLKYLKTDQDQHLVIEVRNKMDSWPBE----- 712
 197 KLDNTIQDAGLYQGVVIEPQEDGTWPRQSLQSKSSTAPSRNFTTSSKPSASPYCSVS 256
 713 MSFTAN-----SSKIDR-----HKVPT---EKGATCLSLNLTGNTCFNNS 748
 257 ASLIANGDSTNNSGMSHSGSVSGSGFSASYNCOEPPSPHIOFGLCGLGNLTGNTCFNNSA 316
 749 IQCVSNTQPTQYFISGRHLYELNRTNPIGMKGMKACVGLVQELWSGTQKNVAPLKL 808
 317 LQCLSNTPATLVEYFLKDEYEAENRDNPLGMKEIAYAYALIKNQWMSGRDTHVAPRMFK 376
 809 WTIKAPRPNFGFQQQDSQELLAFLLDGLHEDLNVRHVKPYVELKDSGDPDWEVAAEAW 868
 377 TVQGRFAPQFSGVQQQDSQELLAFLLDGLHEDLNVRKPKYLEPKDANGRPDAVVAKEAW 436
 869 DNHLRRNRSIVVDLPHGQLRSQVCKTCCHISVRDPNFIPLPMDSYMHEITVIK 928
 437 ENHRLNDSVLVTDTHGLFKSTLVCPCEKAVSVTFDPFCYLTLEPLFKDRIEMEVFLVPA 496
 929 D-GTTPVRYGLRLNMDKTYGLKQLSDCLGNSQIILAEVHGSNI-KNFPQDNQKRL 986
 497 DPQCRPIQYRTVPLMGALISDLCEALSKLSGIAAENMVWTVDVNHRPFKIFQMDGLSHI 556
 987 SVSGFLCAFEIIPVPSIASPTQTDSSSPSTNMTFTLTNGDLPRIPIFIPNGMPTV 1046

Db 715 RSFLADWPELKKRYFDDSAEDFEKHSEVEY--KPPKPEVKLKDCELEFTTKEKLG 772
QY 1251 NEMYYSCKTKHCLATKXLDLWRLPILIIHLKRFQFVNGRWIKSQKIVKPPRESFDP 1310
Db 773 EDPWYCNCKEHQQAATKXLDLWRLPILIIHLKRFQFVNGRWIKSQKIVKPPRESFDP 831
QY 1311 FLVPRDPALCQHKPLTPQGDSEPRILAREVKVYDAQSSAGEEDVLLSKSPSSLSANII 1370
Db 832 --IP----- 833
QY 1371 SSPKSGPSSSRKSGTSCPPSKNSNSPRTLGRSKRLRLPQIGSKNKLSSKENLDAS 1430
Db 834 -----YNPAGP----- 840
QY 1431 KENGAGQICELADALSRGHVGGSOPELVTPQDHEVALANGFLYEACGNGYSNGQLGN 1490
Db 841 -----CR----- 842
QY 1491 HSEEDSTDQREDTRIKPIYNIYALSHSGILGGHYVYIYAKN-ENCKWYCYNDSSCKEL 1549
Db 843 -----YNLIASVSNHYGGMGGGHTAFVKNKDDGKWIYFDDSSVSIA 883
QY 1550 HPEIDTDSAYILFYEQO---CIDYAOFLPKTDGKKA-----DTSSMDEDF 1593
Db 884 CEDQIVSKAAYLVFYORQDTISGTGF--FLDRETKQGASAAATGIPLESDEDSNENDNDI 941
QY 1594 ESD 1596
Db 942 ENE 944

Search completed: February 11, 2004, 15:40:20
Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:32:01 / Search time 53 Seconds
(without alignments)
4803.725 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKESIGFLSYEALRRV.....DTSSMEDPFSDYKVCVLQ 1604

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8563	100.0	1604	23	AAU82715 Amino acid sequenc
2	6826	79.7	1317	23	ABG32160 Novel human ubiqui
3	6807	79.5	1274	23	AAU75578 Human ubiquitin sp
4	5570.5	65.1	1059	24	ABU12111 Human protein mod
5	4822	56.3	1127	22	AAU40477 Human polypeptide
6	4822	56.3	1127	22	AAU40478 Human polypeptide
7	4543	53.1	1085	22	AAU38692 Human polypeptide
8	4531	52.9	1089	19	AAU82396 Human UBP protein
9	4531	52.9	1089	22	ABG23332 Novel human diagno

10	4531	52.9	1089	22	AAU38691 Human polypeptide
11	4433.5	51.8	1125	22	ABG21348 Novel human diagno
12	4328.5	50.5	1388	22	ABG21344 Novel human diagno
13	3138	36.6	582	23	AAE19177 Human protease PR
14	3019.5	35.3	1779	22	ABE1649 Drosophila melanog
15	2711	31.7	539	22	AAU87244 Novel central nerv
16	2509	29.3	1066	22	ABG24025 Novel human diagno
17	2392.5	27.9	462	21	AAU92344 Human cancer assoc
18	1801	21.0	1131	22	ABG21347 Novel human diagno
19	1801	21.0	1142	22	ABG24027 Novel human diagno
20	1750	20.4	401	22	AAU43544 Human polypeptide
21	1406	16.4	805	22	ABG24026 Novel human diagno
22	1140.5	13.3	439	22	ABG21346 Novel human diagno
23	1140.5	13.3	439	22	ABG24024 Novel human diagno
24	1081	12.6	952	21	AAU70014 Human Protease and
25	1053.5	12.3	920	23	ABP68903 Human polypeptide
26	1006	11.7	929	21	ABG31411 Arabidopsis thalia
27	1006	11.7	937	21	ABG31410 Arabidopsis thalia
28	1006	11.7	937	24	ABP59344 AtAG10590 protein
29	938.5	11.0	1024	22	AAU31808 Novel human secret
30	932.5	10.9	835	21	ABG31412 Arabidopsis thalia
31	920.5	10.7	909	21	ABG31701 Arabidopsis thalia
32	920.5	10.7	914	21	ABG31700 Arabidopsis thalia
33	890	10.4	173	22	AAU87545 Novel central nerv
34	890	10.4	173	22	AAU43622 Human polypeptide
35	890	10.4	173	22	AAU19956 Novel human calciu
36	876	10.2	196	24	ABU00140 Human novel polype
37	837.5	9.8	320	22	ABG21345 Novel human diagno
38	787.5	9.2	748	21	ABG31702 Arabidopsis thalia
39	738.5	8.6	1254	22	AAU70760 S cerevisiae apopt
40	713	8.3	1372	22	AAU70881 C albicans apotos
41	665	7.8	1318	22	AAU28169 Novel human secret
42	665	7.8	1373	21	AAU42190 Human ORFX ORF1954
43	647.5	7.6	712	22	ABG00341 Novel human diagno
44	607	7.1	1096	22	ABG60222 Drosophila melanog
45	524	6.1	1123	23	ABU98133 Human PMM incyte

ALIGNMENTS

RESULT 1

AAU82715
ID AAU82715 standard; Protein; 1604 AA.

XX AC AAU82715;

XX DT 23-APR-2002 (first entry)

XX DE Amino acid sequence of novel human protease #14.

XX DE Human; protease; cancer; immune-related disorder; cardiovascular disease;
KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
KW nervous system disorder; sexual dysfunction; pain; mood disorder;
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
KW ocular disease; cytostatic; enzyme.

XX OS Homo sapiens.

XX PN WO200200860-A2.

XX PD 03-JAN-2002.

XX PF 26-JUN-2001; 2001WO-US20171.

XX PR 26-JUN-2000; 2000US-214047P.

XX PA (SUGEN) SUGEN INC.

XX PI Florman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
Chafydzak G;

XX

WPI; 2002-139913/18.
N-PSDB; ABK31757.

Nucleic acids encoding novel human proteases, useful for useful for treating diseases and disorders such as cancers, immune-related diseases and disorders, cardiovascular disease (e.g. restenosis) and inflammatory disorders.

Claim 6; Fig 2F-G; 313pp; English.

The present invention relates to the isolation of novel human proteases, and the nucleic acids encoding them. The sequences of the invention are useful for treating diseases and disorders such as cancers (e.g. breast, colon, lung), immune-related diseases and disorders (e.g. inflammatory diseases and asthma), cardiovascular diseases (e.g. restenosis and coronary thrombosis), brain or neuronal-associated diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory disorders (e.g. rheumatoid arthritis and psoriasis), central or peripheral nervous system diseases, migraines, pain, sexual dysfunction, mood disorders, attention disorders, neurological disorders, hypertension, psychotic disorders, cognitive disorders, hypotension, (e.g. Alzheimer's disease, Parkinson's disease) and dyskinesias. The nucleic acids and polypeptides are also useful for treating viral infections caused by human immunodeficiency virus (HIV), and non-viral infections such as ocular disease (e.g. glaucoma) and macular degeneration. AAU82702-AAU82760 represent the novel human proteases of the invention.

Sequence 1604 AA;

Query Match 100.0%; Score 8563; DB 23; Length 1604;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MGAKSRIGFLSYEERARRVTDVELKRLKDAFRTCTGLSYMGQHCFTREVLGDGVPKV 60
1 MGAKSRIGFLSYEERARRVTDVELKRLKDAFRTCTGLSYMGQHCFTREVLGDGVPKV 60
61 AEVYCSFGTSGKLFNNLIVGLVLTGKDEKAKYIFSLFSSSGNVTREEMRL 120
61 AEVYCSFGTSGKLFNNLIVGLVLTGKDEKAKYIFSLFSSSGNVTREEMRL 120
121 HVVDGKVPDTRKCFSEGEKGVNKEFRNMLFNKDAFTFSRMLLSGGVYVTLTDSDTPT 180
121 HVVDGKVPDTRKCFSEGEKGVNKEFRNMLFNKDAFTFSRMLLSGGVYVTLTDSDTPT 180
181 FYOTLAGVTHLEESDIIDLEKRYLLKKAQSGRTGRFDETFGLVSPPIRPSLSEGLFNAP 240
181 FYOTLAGVTHLEESDIIDLEKRYLLKKAQSGRTGRFDETFGLVSPPIRPSLSEGLFNAP 240
241 DENRDNDHDFKEISCGLSACCRGFLAERQKFCPKVDFDVRDGVLSRVELRDMVVALLEVM 300
241 DENRDNDHDFKEISCGLSACCRGFLAERQKFCPKVDFDVRDGVLSRVELRDMVVALLEVM 300
301 KDNRTDDIPELHNDLSIDIVEGILNADHTTQGHLTIEDYQINSKYNLANEPLNLLFQVC 360
301 KDNRTDDIPELHNDLSIDIVEGILNADHTTQGHLTIEDYQINSKYNLANEPLNLLFQVC 360
361 HIVLGLRAPATPEEGQIIRGWLERSRYGLQAGHNNWFIISMQWQWQKEYVYDANPVVI 420
361 HIVLGLRAPATPEEGQIIRGWLERSRYGLQAGHNNWFIISMQWQWQKEYVYDANPVVI 420
421 EPSVLNGGKYSFGTAAPHEQVEDRIGSLSYNTTEKFSNISTASEASETAGSGFL 480
421 EPSVLNGGKYSFGTAAPHEQVEDRIGSLSYNTTEKFSNISTASEASETAGSGFL 480
481 YSATPGADVCFARQHTSDNNNCLLGANGNILLHNPQKPAIDNQPVTQEPVKATSL 540
481 YSATPGADVCFARQHTSDNNNCLLGANGNILLHNPQKPAIDNQPVTQEPVKATSL 540
541 TLEGGRUKRTQLIHGRDYEMVPEPVWRLYHYWGANLALPRVIRKNSKTDIPELELFFR 600
541 TLEGGRUKRTQLIHGRDYEMVPEPVWRLYHYWGANLALPRVIRKNSKTDIPELELFFR 600

RESULT 2
ABG32160
ID ABG32160 standard; Protein; 1317 AA.

601 YLLFLRQOPATRTQQSNIVWNGNVPSPNAPKRVLAYTTCFSRMOTIKIHEIYLSORLR 660
601 YLLFLRQOPATRTQQSNIVWNGNVPSPNAPKRVLAYTTCFSRMOTIKIHEIYLSORLR 660
661 IKBEDRLMWYNSENVLTLLDDHDKLEYLKTQDEQHLVIEVRNKMDSWPEMSFTANSS 720
661 IKBEDRLMWYNSENVLTLLDDHDKLEYLKTQDEQHLVIEVRNKMDSWPEMSFTANSS 720
721 KIDRHKVPTEKGTGLSNLNTCFMNSSIQCVSNTOPLTQYFISGRHLYELNRTNIGMK 780
721 KIDRHKVPTEKGTGLSNLNTCFMNSSIQCVSNTOPLTQYFISGRHLYELNRTNIGMK 780
781 GHMAKCYGDIIVQELWSTQKNVAPLKLRTWTIAKYAPRFGFQOQDSQELLAFLLDGLHED 840
781 GHMAKCYGDIIVQELWSTQKNVAPLKLRTWTIAKYAPRFGFQOQDSQELLAFLLDGLHED 840
841 LNRVHEKPYVELKSDGRPDWEAAEANDNHLRNRISIVVDLPHGQLRSQVKCTCGHIS 900
841 LNRVHEKPYVELKSDGRPDWEAAEANDNHLRNRISIVVDLPHGQLRSQVKCTCGHIS 900
901 VRFPFNLPLPMDSYMHELTIVIKLDGTTTPRVYGLRLNMDEKYTGKQKLSDLCLN 960
901 VRFPFNLPLPMDSYMHELTIVIKLDGTTTPRVYGLRLNMDEKYTGKQKLSDLCLN 960
961 SEQILLAEVHGSNIKPPQDNQKRVLSVSGFLCAFELPVPVSPISASSPTQTFSSPST 1020
961 SEQILLAEVHGSNIKPPQDNQKRVLSVSGFLCAFELPVPVSPISASSPTQTFSSPST 1020
1021 NEMFTLTNGDLPRPIPIPNMGMPNTVPCGTEKNFTNGMVGMPSLPDSPTFYIIAVH 1080
1021 NEMFTLTNGDLPRPIPIPNMGMPNTVPCGTEKNFTNGMVGMPSLPDSPTFYIIAVH 1080
1081 RKMRTLYFLSSQKRPISLFGMPLIVPCTVHTRKLDYDAVMIQVSRLASPLPQOASN 1140
1081 RKMRTLYFLSSQKRPISLFGMPLIVPCTVHTRKLDYDAVMIQVSRLASPLPQOASN 1140
1141 HAQCDSDSMGYQYPTLTVVQKDGNSCAWCPWTRFCRGCKIDCGEDRAFIAGNAYIAVDWD 1200
1141 HAQCDSDSMGYQYPTLTVVQKDGNSCAWCPWTRFCRGCKIDCGEDRAFIAGNAYIAVDWD 1200
1201 PTALHLYIYQTSQERVDEHESVEQSRRAQAEPLNLSCLRAFTSEELGENEMYCSKCK 1260
1201 PTALHLYIYQTSQERVDEHESVEQSRRAQAEPLNLSCLRAFTSEELGENEMYCSKCK 1260
1261 THCLATKKLDLRLPPLIILHLKRFQVNGRWIKSQKIVKFPRESFDPSPALC 1320
1261 THCLATKKLDLRLPPLIILHLKRFQVNGRWIKSQKIVKFPRESFDPSPALC 1320
1321 QHKPLTPQGDSELPRIAREVKVKVDAQSSAGBEDVLLSKSPSSLSANITSSPKGSPSS 1380
1321 QHKPLTPQGDSELPRIAREVKVKVDAQSSAGBEDVLLSKSPSSLSANITSSPKGSPSS 1380
1381 RKSTCSPSKSNSSPSSPRTLGSKGRLRLPOLGSKNKLSSKENLDASKENGAGQICE 1440
1381 RKSTCSPSKSNSSPSSPRTLGSKGRLRLPOLGSKNKLSSKENLDASKENGAGQICE 1440
1441 LADALSRGHVGGSQPELVTPODHEVALANGFLYEHACNGYSGNGLGNSBEDTDDQ 1500
1441 LADALSRGHVGGSQPELVTPODHEVALANGFLYEHACNGYSGNGLGNSBEDTDDQ 1500
1501 REDTRIPIYNLVYAISSCHSGLTGGGHVTVYAKPNCKWYCYNDSSCKELPDEIDTDSAY 1560
1501 REDTRIPIYNLVYAISSCHSGLTGGGHVTVYAKPNCKWYCYNDSSCKELPDEIDTDSAY 1560
1561 ILFYEQQGIDYVAQFLPKTDGKQVADTSSMDEDFESDYKCYCVLQ 1604
1561 ILFYEQQGIDYVAQFLPKTDGKQVADTSSMDEDFESDYKCYCVLQ 1604

XX	ABG32160;	Best Local Similarity 99.9%; Pred. No. 0;			
AC	29-NOV-2002 (first entry)	Matches 1274; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			
DT	Novel human ubiquitin carboxy-terminal hydrolase 80091.				
DE	Ubiquitin carboxy-terminal hydrolase; haematopoietic disorder; obesity;				
XX	anaemia; drug-induced anaemia; haemolytic anaemia; erythrocytosis;				
KW	chronic renal failure; erythrocytosis; cancer; neutropenia; anorexia;				
KW	granulocytopenia; AIDS; Acquired immunodeficiency syndrome; diabetes;				
KW	myelodysplastic syndrome; Cell proliferation disorder; poliomyelitis;				
KW	cell differentiation disorder; neurological disorder; brain disorder;				
KW	cerebrovascular disease; acute meningitis; intracranial haemorrhage;				
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease; tumour;				
KW	metabolic disorder; cachexia; lipid disorder; angiotonic disorder;				
KW	psoriasis; wound healing; rheumatoid arthritis; Kawasaki syndrome;				
KW	endothelial cell disorder; gene therapy; human.				
XX	Homoc sapiens.				
OS	WO200263031-A2.				
PN	15-AUG-2002.				
XX	07-FEB-2002; 2002WO-US03793.				
PD	07-FEB-2001; 2001US-267054P.				
XX	(MILL-) MILLENNIUM PHARM INC.				
XX	Meyers RE;				
PI	WPI; 2002-627564/67.				
XX	N-PSDB; ABK90784.				
DR	Novel isolated ubiquitin carboxy-terminal hydrolase family polypeptide,				
XX	designated 80091 polypeptides, useful for detecting modulator compounds				
PT	for treating e.g. hematopoietic disorders, cancers and				
PT	granulocytopenias -				
XX	Claim 5; Page 112-113; 123pp; English.				
PS	The invention describes an isolated human ubiquitin carboxy-terminal				
XX	hydrolase family polypeptide, designated 80091 polypeptide (I). (I) is				
CC	useful as reagents or targets in assays applicable to treatment and				
CC	diagnosis of 80091-mediated or -related disorders. Antibodies are useful				
CC	for inhibiting the proliferation or inducing the killing of an 80091				
CC	expressing cell, for treating haematopoietic disorders e.g. anaemia such				
CC	as drug-induced anaemia, haemolytic anaemia, aberrant erythrocytosis,				
CC	anaemia of chronic disease such as chronic renal failure, erythrocytosis				
CC	and/or cancer. Antibodies are also useful for treating or preventing				
CC	neutropenia, granulocytopenia, AIDS, congenital and cyclic neutropenia,				
CC	and myelodysplastic syndromes. 80091 molecules are useful for treating				
CC	and diagnosing haematopoietic disorders, cellular proliferative and/or				
CC	differentiative disorders, neurological or brain disorders,				
CC	cerebrovascular diseases (e.g. acute meningitis, intracranial				
CC	haemorrhage, Alzheimer's disease, Parkinson's disease, Huntington's				
CC	disease and poliomyelitis), metabolic disorders (e.g. obesity, anorexia,				
CC	cachexia, diabetes and lipid disorders), angiogenic disorders (e.g.				
CC	psoriasis, wound healing and complications in diseases such as diabetes				
CC	and rheumatoid arthritis) and endothelial cell disorders (e.g. tumours				
CC	and Kawasaki syndrome). The polynucleotide encoding (I) is useful in gene				
CC	therapy, to express (I), to detect 80091 mRNA or a genetic alteration in				
CC	a 80091 gene, to modulate 80091 activity, in chromosome mapping, to				
CC	identify an individual from a minute biological sample (tissue typing),				
CC	and to aid in forensic identification of the biological sample. This				
CC	is the amino acid sequence of the novel human ubiquitin carboxy-terminal				
CC	hydrolase 80091.				
XX	Sequence 1317 AA;				
XX	Query Match 79.7%; Score 6826; DB 23; Length 1317;				

1063 SAGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPPSKNSPNSPRTLGRSKRL 1122
 1410 RLPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVILGSGQPELVTPQDHEVALA 1469
 1123 RLPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVILGSGQPELVTPQDHEVALA 1182
 1470 NGFLYEACGNGYNGQICNHSSEDSDDQREDTRIKPIYNLYAISCHSILGGGHVY 1529
 1183 NGFLYEACGNGYNGQICNHSSEDSDDQREDTRIKPIYNLYAISCHSILGGGHVY 1242
 1530 YAKNPCKWYCYNDSSCKELHPEIDTDSAYILFYEQQIDYVAQLPXTDGKQWADTSSM 1589
 1243 YAKNPCKWYCYNDSSCKELHPEIDTDSAYILFYEQQIDYVAQLPXTDGKQWADTSSM 1302
 1590 DEFPEDYKCYVLQ 1604
 1303 DEFPEDYKCYVLQ 1317

RESULT 3
 AAU75578
 AAU75578 standard; Protein; 1274 AA.
 AAU75578;
 08-MAY-2002 (first entry)
 Human ubiquitin specific protease 10 (USP10).
 Ubiquitin specific protease 10; USP10; human; testis.
 Homo sapiens.
 CN1319662-A.
 31-OCT-2001.
 11-APR-2001; 2001CN-0113514.
 11-APR-2001; 2001CN-0113514.
 (UYNA-) UNIV NANJING MEDICAL.
 Sha J, Zhou Z, Li J;
 WPI; 2002-123068/17.
 N-PSDB; ABK14799.
 Human testicular ubiquitin specific proteinase 10 gene coding protein -
 Claim 1; Page 1(Claims); 7pp; Chinese.
 The invention relates to a human testis ubiquitin specific protease 10, (USP10) gene, and protein. The USP10 gene can be used to prepare a fusion protein, which can be used to prepare mono-/polyclonal antibody. The gene can be also be used for preparing testis specific function gene expression chip, in gene therapy for curing USP10-related diseases. The present sequence represents the amino acid sequence of human USP10.
 Sequence 1274 AA;
 Query Match 79.5%; Score 6807; DB 23; Length 1274;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 331 MGLTLEDYQIWSVKNVLANEFANLIFQVCHIVGLRPAATPEEGQIRGWLERESRYGL 390
 1 MGLTLEDYQIWSVKNVLANEFANLIFQVCHIVGLRPAATPEEGQIRGWLERESRYGL 60
 391 QAGHNWFIISMQWQWQWKEYVVDANPVIEPSSVLNGKYSFGTAHPMEQVEDRIGSS 450
 61 QAGHNWFIISMQWQWQWKEYVVDANPVIEPSSVLNGKYSFGTAHPMEQVEDRIGSS 120

QY 451 LSVYNTTEKFSNISTASEASETAGSGFLYSATPGADVCFARQHNNTSDNNQCILGANG 510
 DB 121 LSVYNTTEKFSNISTASEASETAGSGFLYSATPGADVCFARQHNNTSDNNQCILGANG 180
 QY 511 NILHLNPOKPGAIHQPLVTOEPVKATSLTLEGGRLKRTPOLIHGRDYEMPEPVRAL 570
 DB 181 NILHLNPOKPGAIHQPLVTOEPVKATSLTLEGGRLKRTPOLIHGRDYEMPEPVRAL 240
 QY 571 YHWYGANLALPREVIKNSKTDIPELFPYRLLFRQQPATRTOQSNIWVNMGNVSPNA 630
 DB 241 YHWYGANLALPREVIKNSKTDIPELFPYRLLFRQQPATRTOQSNIWVNMGNVSPNA 300
 QY 631 PLKRVLAYTCFSRMTIKIIEHYLSQRIRIKEEDMRWLWYNSENYLTLLDDEHKLVL 690
 DB 301 PLKRVLAYTCFSRMTIKIIEHYLSQRIRIKEEDMRWLWYNSENYLTLLDDEHKLVL 360
 QY 691 KIODEQLVIEVRNKMSPBEMSFIANSKIDRHKVPTKGGATGLSNLGNTCFNMSSIQ 750
 DB 361 KIODEQLVIEVRNKMSPBEMSFIANSKIDRHKVPTKGGATGLSNLGNTCFNMSSIQ 420
 QY 751 CVSNTOPLTQYFISGRHLYELNFTPIGMKGHWAKCYGDLVOELMSGTQKNVAPLKLRT 810
 DB 421 CVSNTOPLTQYFISGRHLYELNFTPIGMKGHWAKCYGDLVOELMSGTQKNVAPLKLRT 480
 QY 811 IAKYAPRFNGFQQDSQELLAFLLDGLHEDLNARVHEKPYVELKSDGRPDWEAAEAWN 870
 DB 481 IAKYAPRFNGFQQDSQELLAFLLDGLHEDLNARVHEKPYVELKSDGRPDWEAAEAWN 540
 QY 871 HLRRNRSIVVDLPHGQIRSQVKCTGCHISVAPDPFNFLSLPLPMDSYMHELTIVIKLDG 930
 DB 541 HLRRNRSIVVDLPHGQIRSQVKCTGCHISVAPDPFNFLSLPLPMDSYMHELTIVIKLDG 600
 QY 931 TTPVRYGLRLNMDKVTGLKQSLDCLNSEQIILAEVHGSNIKFNFPQDNOKVRLSVSG 990
 DB 601 TTPVRYGLRLNMDKVTGLKQSLDCLNSEQIILAEVHGSNIKFNFPQDNOKVRLSVSG 660
 QY 991 FLCAFEIPVPSISASPTQDTFSSSPSTNEMFTLTNGDLPRPIFINGMPTNVVPCG 1050
 DB 661 FLCAFEIPVPSISASPTQDTFSSSPSTNEMFTLTNGDLPRPIFINGMPTNVVPCG 720
 QY 1051 TEKNFTNGMVGNGHMPSLPDSPTFGYIIAHRKMRTELFLSSOKNRPSLFGMLIVPCT 1110
 DB 721 TEKNFTNGMVGNGHMPSLPDSPTFGYIIAHRKMRTELFLSSOKNRPSLFGMLIVPCT 780
 QY 1111 VHTRKKDLYDAVMIQVSRLASPLPQEAASNAHDCDDSMGYQYPTFLRVVQKGNSCAWC 1170
 DB 781 VHTRKKDLYDAVMIQVSRLASPLPQEAASNAHDCDDSMGYQYPTFLRVVQKGNSCAWC 840
 QY 1171 PWYRRCGCKIDCGEDRAFIQNAIYAVDMDPTALHLYOTSQBRVVDEHESVQSRRAQ 1230
 DB 841 PWYRRCGCKIDCGEDRAFIQNAIYAVDMDPTALHLYOTSQBRVVDEHESVQSRRAQ 900
 QY 1231 EPINLDSCLRAFTSEELGENEMYCKKTHCLATKCLDLMLRPLPILIHKLRFVNG 1290
 DB 901 EPINLDSCLRAFTSEELGENEMYCKKTHCLATKCLDLMLRPLPILIHKLRFVNG 960
 QY 1291 RWIKSQIVKVPRESPPSAFLVPRPALCOHQPLTPQGDSELPILAREVKVQAQSS 1350
 DB 961 RWIKSQIVKVPRESPPSAFLVPRPALCOHQPLTPQGDSELPILAREVKVQAQSS 1020
 QY 1351 AGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPPSKNSPNSPRTLGRSKRL 1410
 DB 1021 AGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPPSKNSPNSPRTLGRSKRL 1080
 QY 1411 LPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVILGSGQPELVTPQDHEVALA 1470
 DB 1081 LPQIGSKNKLSSSKENLDASKENGAGQICELADALSRGHVILGSGQPELVTPQDHEVALA 1140
 QY 1471 GFLYEACGNGYNGQICNHSSEDSDDQREDTRIKPIYNLYAISCHSILGGGHVY 1530
 DB 1141 GFLYEACGNGYNGQICNHSSEDSDDQREDTRIKPIYNLYAISCHSILGGGHVY 1200
 QY 1531 AKNPCKWYCYNDSSCKELHPEIDTDSAYILFYEQQIDYVAQLPXTDGKQWADTSSMD 1590

|||||
Db 1201 AKNPNCKWYCYNDSSCKELHPDEIDTBSAYILFVEQOIGIDVAQLPXTDGGKQADTSSMD 1260
Qy 1591 EDFESDYKKYCVLQ 1604
Db 1261 EDFESDYKKYCVLQ 1274
RESULT 4
ABU12111
ID ABU12111 standard; Protein; 1059 AA.
XX
AC ABU12111;
XX
DT 17-FEB-2003 (first entry)
XX
DE Human protein modification and maintenance molecule (PMOD) #8.
XX
KW Human; protein modification and maintenance molecule; PMOD;
KW gastrointestinal disorder; anorexia; nausea; cardiovascular disorder;
KW angina pectoris; myocardial infarction; autoimmune disorder; anaemia;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS; gout;
KW developmental disorder; epilepsy; epithelial disorder; vitiligo; eczema;
KW neurological disorder; Huntington's disease; Parkinson's disease;
KW reproductive disorder; infertility; impotence; incyte 7485451CD1.
XX
OS Homo sapiens.
XX
PN WO200281636-A2.
XX
PD 17-OCT-2002.
XX
PF 05-APR-2002; 2002WO-US10812.
XX
PR 05-APR-2001; 2001US-282282P.
PR 13-APR-2001; 2001US-283782P.
PR 18-APR-2001; 2001US-284823P.
PR 27-APR-2001; 2001US-287264P.
PR 04-MAY-2001; 2001US-288662P.
PR 11-MAY-2001; 2001US-290383P.
PR 15-JUN-2001; 2001US-298348P.
PR 25-JAN-2002; 2002US-351928P.
PR 25-FEB-2002; 2002US-359903P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Deleage AM, Swarnakar A, Hafalia AJA, Duggan BM;
PI Warren BA, Emerling BM, Arvizu CS, Honchell CD, Kallick DA;
PI Lu DAM, Lee EA, Yue H, Forsythe J, Rankumar J, Griffin JA;
PI Li JX, Thangavelu K, Baughn MR, Yao MG, Sanjanwala MM, Wallia NK;
PI Burford N, Lal PG, Becha SD, Lee SY, Elliott VS, Luo W, Lu Y,
PI Wang YE;
XX
DR WPI; 2003-058519/05.
DR N-PSDB; ABX56506.
XX
PT Novel isolated human protein modification and maintenance molecules
PT useful for diagnosing, treating, preventing Alzheimer's disease,
PT epilepsy, gout, impotence, anaemia, angina pectoris, vitiligo, anorexia,
PT nausea
XX
PS Claim 1; Page 147-149; 184pp; English.
XX
CC The invention relates to human protein modification and maintenance
CC molecules (PMOD) and to polynucleotides encoding them. The polypeptides
CC of gastrointestinal disorders (e.g. anorexia, nausea, treatment and prevention
CC of disorders (e.g. angina pectoris, myocardial infarction,
CC autoimmune/inflammatory disorders, e.g. acquired immunodeficiency
CC syndrome (AIDS), gout), developmental disorders (e.g. anaemia, epilepsy),
CC epithelial disorders (e.g. vitiligo, eczema), neurological disorders
CC (e.g. Huntington's disease, Parkinson's disease) and reproductive
CC disorders (e.g. infertility, impotence). The polypeptides may be used in

assays for detecting the presence of the associated disorders. The
polynucleotides are useful for detecting upstream sequences such as
promoters and regulatory elements, for creating knock out or knock in
humanised animals or transgenic animals to model human diseases and in
somatic or germline gene therapy for treating the disorders. Sequences
ABU12104-ABU12120 represent human protein modification and maintenance
molecules of the invention.
XX
SQ Sequence 1059 AA;
Query Match 65.1%; Score 5570.5; DB 24; Length 1059;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1044; Conservative 0; Mismatches 0; Indels 15; Gaps 2;
Qy 561 MYPEPWALYHWYGANLALPRVFNKSKTDIPELELPYLLFLRQOPATRTQOSNIWV 620
Db 1 MYPEPWALYHWYGANLALPRVFNKSKTDIPELELPYLLFLRQOPATRTQOSNIWV 60
Qy 621 NM-----GNVPSNAPLKRVLAYTGTFSRMQTIKEIHEYLSQRLRIKEEDM 666
Db 61 NMGMSLRMFPQHLPRGNVPSNAPLKRVLAYTGTFSRMQTIKEIHEYLSQRLRIKEEDM 120
Qy 667 RLWLYNSENLYLLDDEHKLKLYLKIODEQHLVIVRNKMSWPEMSFIANSKIDRHK 726
Db 121 RLWLYNSENLYLLDDEHKLKLYLKIODEQHLVIVRNKMSWPEMSFIANSKIDRHK 180
Qy 727 VPTKGATGLSNLNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRPINPMKGMMAK 786
Db 181 VPTKGATGLSNLNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRPINPMKGMMAK 240
Qy 787 YGDLVQELWSGTQKNVAPLKLWTTAKYAPRNGQQDSQELLAFLLDGLHEDLNRVHE 846
Db 241 YGDLVQELWSGTQKNVAPLKLWTTAKYAPRNGQQDSQELLAFLLDGLHEDLNRVHE 300
Qy 847 KPVELKDSGDRPDWEVAEAWNDHLRNRSLVVDLPHGQLRSQVKCTCGHISVRFDFP 906
Db 301 KPVELKDSGDRPDWEVAEAWNDHLRNRSLVVDLPHGQLRSQVKCTCGHISVRFDFP 360
Qy 907 NFLSLPLMDSYMHLEITVIKLDGTPVRYGLRLNMDEKYTGKQLSDLCGLNSEQILL 966
Db 361 NFLSLPLMDSYMHLEITVIKLDGTPVRYGLRLNMDEKYTGKQLSDLCGLNSEQILL 420
Qy 967 AEVHGSNIKFPQDNQKRLSVSGFLCAFEIPVPSPISSASPTQDFSSSSTNEMFTL 1026
Db 421 AEVHGSNIKFPQDNQKRLSVSGFLCAFEIPVPSPISSASPTQDFSSSSTNEMFTL 480
Qy 1027 TTNGDLPRFIPFNGMNPNTVPCGTEKNTGVMNGHMPSPDPSPTGYIIAVHKKMRT 1086
Db 481 TTNGDLPRFIPFNGMNPNTVPCGTEKNTGVMNGHMPSPDPSPTGYIIAVHKKMRT 540
Qy 1087 ELYFLSSQKRPFLGMPPLIVPCTVHTRKOLYDAVWIOVSRSLASPLPPQEASNAHQDCD 1146
Db 541 ELYFLSSQKRPFLGMPPLIVPCTVHTRKOLYDAVWIOVSRSLASPLPPQEASNAHQDCD 600
Qy 1147 DSMGYQYPTLWVQKGNSCAWCPWYRFCRCCKIDCGEDRAFIQNAIYAVDWDPTALHL 1206
Db 601 DSMGYQYPTLWVQKGNSCAWCPWYRFCRCCKIDCGEDRAFIQNAIYAVDWDPTALHL 660
Qy 1207 RYQTSQERVDEHESVQSRRAQAEPIINLDSCLRAFTSEELGENEMYCSKCKTHCLAT 1266
Db 661 RYQTSQERVDEHESVQSRRAQAEPIINLDSCLRAFTSEELGENEMYCSKCKTHCLAT 720
Qy 1267 KKLDELWLPILIIHLKRPQVNGWIKSQIKVFPRESFDSALVPRDPAALCOHKPLT 1326
Db 721 KKLDELWLPILIIHLKRPQVNGWIKSQIKVFPRESFDSALVPRDPAALCOHKPLT 780
Qy 1327 PQGDELSEPRILAREVKKYVDAQSSAGEEDVLLSKSPSSLSANIISSPKGSPPSSSRKSGTS 1386
Db 781 PQGDELSEPRILAREVKKYVDAQSSAGEEDVLLSKSPSSLSANIISSPKGSPPSSSRKSGTS 840
Qy 1387 CPSSKNSSPNSSPRLTGRSKGRRLRPLQIGSKNKLSSSKENLDASKENGAGQICELADALS 1446
Db 841 CPSSKNSSPNSSPRLTGRSKGRRLRPLQIGSKNKLSSSKENLDASKENGAGQICELADALS 900

QY 1447 RGHVLG-GSQPELVTPQDHEVALANGFLYHEACGNGYNGQIGNHSEEDSTDQREDTR 1505
 DB 901 RGHVLGVSQPELVTPQDHEVALANGFLYHEACGNGYNGQIGNHSEEDSTDQREDTR 960
 QY 1506 IKPIYNLYAISCHSGILGGHYVTYAKNPCKWYCYNDSCKELHPDEIDTDSAYILFYE 1565
 DB 961 IKPIYNLYAISCHSGILGGHYVTYAKNPCKWYCYNDSCKELHPDEIDTDSAYILFYE 1020
 QY 1566 QQGIDYAQFLPKTDGKWADTSSNDEDFESDYKICYVLQ 1604
 DB 1021 QQGIDYAQFLPKTDGKWADTSSNDEDFESDYKICYVLQ 1059

RESULT 5

AAM40477

ID AAM40477 standard; Protein; 1127 AA.

XX AC AAM40477;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 5408.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Sky-Brager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX XX 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 13-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX XX (HYSE-) HYSEQ INC.

XX PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX XX WPI; 2001-442253/47.

XX DR N-PSDB; AAI59633.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders
 XX PT such as central nervous system injuries -

XX PS Example 2; SEQ ID NO 5408; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAI42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Sky-Brager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 1127 AA;

Query Match 56.3%; Score 4822; DB 22; Length 1127;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 901; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 701 EVRNKDMSPREMSFIANSKKIDRHKVPTKGTAGLNLGNTCFMNSSIQCVSNTPQLTQ 760

DB 220 EVRNKDMSPREMSFIANSKKIDRHKVPTKGTAGLNLGNTCFMNSSIQCVSNTPQLTQ 279

QY 761 YFISGRHLYELNRTNPIGMKGHMAKCYGLVQELWSGTQKNVAPLKLRTWTIAKYPFNG 820

DB 280 YFISGRHLYELNRTNPIGMKGHMAKCYGLVQELWSGTQKNVAPLKLRTWTIAKYPFNG 339

QY 821 FQOQDSQELLAFLLDGLHEDLNVRHKKPYVELKDSGRPDWEVAAEADNHLNRRTIV 880

DB 340 FQOQDSQELLAFLLDGLHEDLNVRHKKPYVELKDSGRPDWEVAAEADNHLNRRTIV 399

QY 881 DLFHQQLRSQVKCKTCGHSIVRFPFNFSLPLPMDSYMHLEITVIKLDGTPVRYGLRL 940

DB 400 DLFHQQLRSQVKCKTCGHSIVRFPFNFSLPLPMDSYMHLEITVIKLDGTPVRYGLRL 459

QY 941 NMDEKYGKQLSDLCGLNSEQILLARVHGSNKNPQDNQKVRLLSVSGFLCAFEIPLVP 1000

DB 460 NMDEKYGKQLSDLCGLNSEQILLARVHGSNKNPQDNQKVRLLSVSGFLCAFEIPLVP 519

QY 1001 VSPISASPTQTDPSSTPTNEMFTLTNGDLPRPIPIPNMGMPNTVPCGTEKNTNGMV 1060

DB 520 VSPISASPTQTDPSSTPTNEMFTLTNGDLPRPIPIPNMGMPNTVPCGTEKNTNGMV 579

QY 1061 NGHMPSLPDSPTGYIIAIVHKKMRTELYFLSSQKNRPSLFGMPLIIVCTVHTKKDLYD 1120

DB 580 NGHMPSLPDSPTGYIIAIVHKKMRTELYFLSSQKNRPSLFGMPLIIVCTVHTKKDLYD 639

QY 1121 AVWIOVSRSLASPLPQEAASNAQDCDDSMGYQYFTLRVQKDGNSCAWCWYFCRGCK 1180

DB 640 AVWIOVSRSLASPLPQEAASNAQDCDDSMGYQYFTLRVQKDGNSCAWCWYFCRGCK 699

QY 1181 IDCGEDRAFIAGNAYIADVNDPTALHLYRYQTSQERVVDEHESVEQSRAQAQPEINLDSCLR 1240

DB 700 IDCGEDRAFIAGNAYIADVNDPTALHLYRYQTSQERVVDEHESVEQSRAQAQPEINLDSCLR 759

QY 1241 APTSEELGENEMYCSKCKTHCLATKCLDLWRLLPILIIHLKRFQFVNGRWIKSQIKV 1300

DB 760 APTSEELGENEMYCSKCKTHCLATKCLDLWRLLPILIIHLKRFQFVNGRWIKSQIKV 819

QY 1301 PPRSFDPDSAFIVPRDPALCOHKPLTPQDSELSRPRILAREVKVDAQSSAGEDVLLSK 1360

DB 820 PPRSFDPDSAFIVPRDPALCOHKPLTPQDSELSRPRILAREVKVDAQSSAGEDVLLSK 879

QY 1361 SPSSLSANIISSPKGSPSSSRKSGTSCPSKNSPNSSPRTLGRSKGRLLRPLQIGSKNKL 1420

DB 880 SPSSLSANIISSPKGSPSSSRKSGTSCPSKNSPNSSPRTLGRSKGRLLRPLQIGSKNKL 939

QY 1421 SSSKENLDASKENGAGQICELADALSRLHVLGGSOPELVTPQDHEVALANGFLYEHEA-- 1478

DB 940 SSSKENLDASKENGAGQICELADALSRLHVLGGSOPELVTPQDHEVALANGFLYEHEACG 999

QY 1479 --CGNGYNGQIGNHSEEDSTDQREDTRIKPIYNLYAISCHSGILGGHYVTYAKNPNC 1536

DB 1000 NGCGNGYNGQIGNHSEEDSTDQREDTRIKPIYNLYAISCHSGILGGHYVTYAKNPNC 1059

QY 1537 KYCYNDSSCKELHPDEIDTDSAYILFYEQGIDYAFPLPKTDGKKWADTSSMDEDFESD 1596

DB 1060 KYCYNDSSCKELHPDEIDTDSAYILFYEQGIDYAFPLPKTDGKKWADTSSMDEDFESD 1119

QY 1597 YKCYCVLQ 1604

DT	22-OCT-2001	(first entry)	
XX	Human polypeptide SEQ ID NO 1837.		
DE	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
XX	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX	Homo sapiens.		
OS	Human sapiens.		
XX	WO200153312-A1.		
PN	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX	WPI; 2001-442253/47.		
DR	N-PSDB; AA157848.		
XX	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX	Example 3; SEQ ID NO 1837; 10078pp; English.		
BS	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AA138642-AA142213) with neotropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX	Sequence 1085 AA;		
XX	Query Match 53.1%; Score 4543; DB 22; Length 1085;		
XX	Best Local Similarity 94.1%; Pred. No. 0;		
XX	Matches 851; Conservative 22; Mismatches 31; Indels 0; Gaps 0;		
QY	701 EVRNKMSWPEEMSFANSSKIDRHKVPTKGTGLNLGNTCFMNSSIQCVSNTQPLTQ 760		
DB	182 EVHNKMSWPEEMSFANSSKIDRHKVPTKGTGLNLGNTCFMNSSIQCVSNTQPLTQ 241		
QY	761 YFISGRHLYELNRTNPTGMKGMAKCYGDLVQELWSGTQKVAFLKLRWTIAKAPRNG 820		
DB	242 YFISGRHLYELNRTNPTGMKGMAKCYGDLVQELWSGTQKVAFLKLRWTIAKAPKFDG 301		
QY	821 FQQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAAEAWDNHLNRNSIVV 880		
DB	302 FQQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSGRPDWEVAAEAWDNHLNRNSIVV 361		
QY	881 DLPHGQRSQVKCKTCGHSIVRDPNPNFLSLPLPMSYHLEITVILKLDGTTTTPRVGLRL 940		
DB	362 DLPHGQRSQVKCKTCGHSIVRDPNPNFLSLPLPMSYHLEITVILKLDGTTTTPRVGLRL 421		
QY	941 NMDEKTYGLKKQLSDLCGLNSEQILLAEVHGSNIKPFQDNQKVRSLVSGFLCAFEIPVP 1000		
DB	422 NMDEKTYGLKKQLSDLCGLNSEQILLAEVHGSNIKPFQDNQKVRSLVSGFLCAFEIPVP 481		
QY	1001 VSPISASSPTQTFSSSPSTNEMFTLTNGDLPRPIPIPNGMNTVVPCTGTEKNTFNGMV 1060		
DB	482 SSPISASSPTQIDFSSSPSTNGMFTLTNGDLPRPIPIPNGMNTVVPCTGTEKNTFNGMV 541		
QY	1061 NGHMPSLPDSPTFTGYIIAVHRKWMRTLEYFLSKQKRPSLFGMPLIIVPCTVHTRKKDLYD 1120		
DB	542 NGHMPSLPDSPTFTGYIIAVHRKWMRTLEYFLSKQKRPSLFGMPLIIVPCTVHTRKKDLYD 601		
QY	1121 AVTIQVSRLASPLPPQASNAHQDCCDSMGYQYPTLTVVQKDGNSCAMCPWTRFRGCK 1180		
DB	602 AVTIQVSRLASPLPPQASNAHQDCCDSMGYQYPTLTVVQKDGNSCAMCPWTRFRGCK 661		
QY	1181 IDCEDRAFTGNAYIAVDWPTALHLRYQTSQSRVVDHESVQSRRAAQAEPLNLDCLR 1240		
DB	662 IDCEDRAFTGNAYIAVDWPTALHLRYQTSQSRVVDHESVQSRRAAQAEPLNLDCLR 721		
QY	1241 AFTSEEBELGENEMYCYCKKTHCLATKLDLWLPPLIILHLKRFQVNGRWIKSQKIVK 1300		
DB	722 AFTSEEBELGENEMYCYCKKTHCLATKLDLWLPPLIILHLKRFQVNGRWIKSQKIVK 781		
QY	1301 FPRSFDPSPAFVPRDPAALCOHKPLTPQGDSELPRIAREVKKVDAQSSAGEDVLLSK 1360		
DB	782 FPRSFDPSPAFVPRDPAALCOHKPLTPQGDSELPRIAREVKKVDAQSSAGEDVLLSK 841		
QY	1361 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGSRKGLRLPQIGSKNKL 1420		
DB	842 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGSRKGLRLPQIGSKNKL 901		
QY	1421 SSSKENLDASKENGAGQICELADALSRGHVLGSGOPELVTPQDHEVALANGFLYEACG 1480		
DB	902 SSSKENLDASKENGAGQICELADALSRGHVLGSGOPELVTPQDHEVALANGFLYEACG 961		
QY	1481 NGYNGQLGNHSEEDSTDQREDTRIKPIYNLYAISCHSILGGCHVYVYAKPNCKWYC 1540		
DB	962 NGYNGQLGNHSEEDSTDQREDTRIKPIYNLYAISCHSILGGCHVYVYAKPNCKWYC 1021		
QY	1541 YNDSCKELHPDEIDTDSAYILFVEQGGIDYAPLPKTDGKKWADTSSMDEPESDYKY 1600		
DB	1022 YNDSCKELHPDEIDTDSAYILFVEQGGIDYAPLPKTDGKKWADTSSMDEPESDYKY 1081		
QY	1601 CVLQ 1604		
DB	1082 SMLQ 1085		
XX	RESULT 8		
XX	AAW82396		
XX	ID AAW82396 standard; Protein; 1089 AA.		
XX	AC AAW82396;		
XX	DT 23-FEB-1999 (first entry)		
XX	Human UB protein #2.		
XX	Ubiquitin isopeptidase; UBP; intracellular protein; degradation; therapy;		
XX	cell growth; diagnosis; tumour; proliferative disease.		
XX	Homo sapiens.		
PN	WO9848020-A2.		

XX PD 29-OCT-1998.
XX PF 10-APR-1998; 98WO-IT00084.
XX PR 18-APR-1997; 97IT-MI00905.
XX PA (EUON-) IST EURO DI ONCOLOGIA SRL.
XX PI Bosari S, Capra M, Di Fiore PP, Draetta GF, Goubin P,
XX PI Matoskova B, Naviglio S, Soncini C;
XX DR WPI; 1998-609991/51.
XX PS N-PSDB; AAV73472.
XX XX
XX PT New ubiquitin isopeptidases that control cell proliferation - useful
XX PT in the diagnosis and treatment of tumours and other proliferative
XX PT conditions
XX PS Claim 1; Page 44-48; 75pp; English.
XX XX
XX CC This sequence represents a novel human ubiquitin isopeptidase, UBP.
XX CC Ubiquitin isopeptidases are involved in the ubiquitin pathway which
XX CC controls the degradation of a large number of intracellular proteins,
XX CC many of which play key roles in controlling cell growth. Antibodies
XX CC raised against this protein may be used in the diagnosis of a tumour or
XX CC another proliferative disease or for therapeutic or diagnostic purposes.
XX XX
XX SQ Sequence 1089 AA;
Query Match 52.9%; Score 4531; DB 19; Length 1089;
Best Local Similarity 93.7%; Pred. No. 0;
Matches 851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;
QY 701 EVRNKMSWPEEKSFIANSKIDRHKVPTKKGATGLNLGNTCFMNSSIQCVNTQPLTQ 760
DB 182 EVRNKMSWPEEKSFIANSKIDRHKVPTKKGATGLNLGNTCFMNSSIQCVNTQPLTQ 241
QY 761 YFISGRHLYELNRTNPIMKGHMAKCYDILVQLWSGTQKVAFLKRLMTIAKAPRFG 820
DB 242 YFISGRHLYELNRTNPIMKGHMAKCYDILVQLWSGTQKVAFLKRLMTIAKAPRFG 301
QY 821 FQODSDELLAFLLDGLHEDLNVEHKPYVELKDSGRPDWEVAEAWNDHLRNRISIV 880
DB 302 FQODSDELLAFLLDGLHEDLNVEHKPYVELKDSGRPDWEVAEAWNDHLRNRISIV 361
QY 881 DLPHGLRSQVKCTCGHISVRFPDFNFLSLPLPMDSYMDEITVIKLDGTTVPVYGLRL 940
DB 362 DLPHGLRSQVKCTCGHISVRFPDFNFLSLPLPMDSYMDEITVIKLDGTTVPVYGLRL 421
QY 941 NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKFPQDNQKRLSVSGFLCAFEIPVP 1000
DB 422 NMDEKTYGLKQLSDLCGLNSEQILLAEVHGSNIKFPQDNQKRLSVSGFLCAFEIPVP 481
QY 1001 VSPISASSPTQDFSSSPSTNEMFTLTNGDLPRPIFINGMENTVVPCKTEKFTNGMV 1060
DB 482 SPSISASSPTQDFSSSPSTNEMFTLTNGDLPRPIFINGMENTVVPCKTEKFTNGMV 541
QY 1061 NGHMPSLPDPTGYIIVHVKMRTELVELSLQKRPRLFGMLPIVPCVTHTRKDLVD 1120
DB 542 NGHMPSLPDPTGYIIVHVKMRTELVELSLQKRPRLFGMLPIVPCVTHTRKDLVD 601
QY 1121 AVMTQVSRLASPLPPQASNAHDCDSMGVQYFPTLRVQKDGNSCANCWPVRCGCK 1180
DB 602 AVMTQVSRLASPLPPQASNAHDCDSMGVQYFPTLRVQKDGNSCANCWPVRCGCK 661
QY 1181 IDCEDRAFTGNAYIAVMDPTALHLYQTSQRRVVDHESVEQSRAQAEPINLDSCLR 1240
DB 662 IDCEDRAFTGNAYIAVMDPTALHLYQTSQRRVVDHESVEQSRAQAEPINLDSCLR 721
QY 1241 AFTSEELGENEMYCKCKTHCLAKKLDLWRLPILIIHLKRFQVNGRWIKSQKIVR 1300
DB 722 AFTSEELGENEMYCKCKTHCLAKKLDLWRLPILIIHLKRFQVNDWIKSQKIVR 781

QY 1301 FPRESDPSAFILVPRDPALCOHKPLTPQGBELSEPRILAREVKKYDAQSSAGEEDVLISK 1360
DB 782 FLRESFDPFAFLVPRDPALCOHKPLTPQGBELSEPRILAREVKKYDAQSSAGKEDMLISK 841
QY 1361 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGRSKGLRLPQIGSKNKL 1420
DB 842 SPSSLSANISSPKGSPSSSRKSGTSCFSSKNSSPNSPRTLGRSKGLRLPQIGSKNKP 901
QY 1421 SSSKENLDASKENGAGQICELADALSRGHVGLGGQPELVTPQDHEVALANGFLYEHAAG 1480
DB 902 SSSKENLDASKENGAGQICELADALSRGHVGLGGQPELVTPQDHEVALANGFLYEHAAG 961
QY 1481 N----GYSNGQLGNHSEEDSTDDQEDTRIKPIYNYLYAISCHSILGCGHYVYAKNPNC 1536
DB 962 NCGGDSYNGQLGNHSEEDSTDDQEDTRIKPIYNYLYAISCHSILGCGHYVYAKNPNC 1021
QY 1537 KWCYNDSSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTIDGKMDATSSMDEDFESD 1596
DB 1022 KWCYNDSSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTIDGKMDATSSMDEDFESD 1081
QY 1597 YKCYVQLQ 1604
DB 1082 YKCYVQLQ 1089
RESULT 9
ABG23332
ID ABG23332 standard; Protein; 1089 AA.
XX AC ABG23332;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #23323.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS87519.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX PS Claim 20; SEQ ID No 53691; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as

2Y 1481 N-----GYSNGQLGNHSEEDSTDDQREDTRIKPIYNLYAISCHGILGGHYVITYAKNPNC 1536

CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX

Sequence 1089 AA;
Query Match 52.9%; Score 4531; DB 22; Length 1089;
Best Local Similarity 93.7%; Pred. No. 0;
Matches 851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;
QY 701 EVRNKMSWPMSPIANSKIDRHKVPTEKGATGLNLGNTCFMNSSIOCVSNTPQLTQ 760
Db 182 EVRNKMSWPMSPIANSKIDRHKVPTEKGATGLNLGNTCFMNSSIOCVSNTPQLTQ 241
QY 761 YFISGRHLYELNRTNPIGMKGMAKCYGDLVQELWSGTOKNVAFLKLRWTIAKYAPFNG 820
Db 242 YFISGRHLYELNRTNPIGMKGMAKCYGDLVQELWSGTOKNVAFLKLRWTIAKYAPFNG 301
QY 821 FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSQGRPDWEVAEAWDNLHNRNSIIV 880
Db 302 FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKDSQGRPDWEVAEAWDNLHNRNSIIV 361
QY 881 DLFGQLRSQVKCTCGHISVREDFNPLSLPLMDSVMHLEITVVKLDGTTTPRYGLRL 940
Db 362 DLFGQLRSQVKCTCGHISVREDFNPLSLPLMDSVMHLEITVVKLDGTTTPRYGLRL 421
QY 941 NWDEKYTLKKQLSDLCGLNSEQIILAEVHGSNIKFPQKNQKRLSVSGFLCAFEIPVP 1000
Db 422 NWDEKYTLKKQLSDLCGLNSEQIILAEVHGSNIKFPQKNQKRLSVSGFLCAFEIPVP 481
QY 1001 VSPISASPTQTDSSSPSTNEMTLTNGDLPRPIFPNGMPNVTVPCTGKXNFNGMV 1060
Db 482 SSPISASPTQTDSSSPSTNEMTLTNGDLPRPIFPNGMPNVTVPCTGKXNFNGMV 541
QY 1061 NGRMPSLPDSFPTGYIIVHVKMRTLEYFLSSOKNRPFLFGMLIIVPCVTHRKDLYD 1120
Db 542 NGRMPSLPDSFPTGYIIVHVKMRTLEYFLSSOKNRPFLFGMLIIVPCVTHRKDLYD 601
QY 1121 AVTIQVRLASPLPQBRASNAHQDCDSMGQYPTLTVVQKNSCAMCPWYRFCRGCK 1180
Db 602 AVTIQVRLASPLPQBRASNAHQDCDSMGQYPTLTVVQKNSCAMCPWYRFCRGCK 661
QY 1181 IDCEDRAFIGNAVIADWDPTALHLAYQTSOERVDHESVQSRQAEPINLDSCLR 1240
Db 562 IDCEDRAFIGNAVIADWDPTALHLAYQTSOERVDHESVQSRQAEPINLDSCLR 721
QY 1241 AFTSEELGENEMYCKCKTHCLATKCLDLRLPFIILHKKFQFVNGRWIKSQIVK 1300
Db 722 AFTSEELGENEMYCKCKTHCLATKCLDLRLPFIILHKKFQFVNGRWIKSQIVK 781
QY 1301 FPRSPDPSAFLVRDPALCOHKPLTQDELSPRILAREVKVDQAQSGAEDVLLSK 1360
Db 782 FLRESFDSAFLVRDPALCOHKPLTQDELSPRILAREVKVDQAQSGAEDVLLSK 841
QY 1361 SPSSLSANISSPKGSPSSSKGTSQSPSSKNSSPNSPRTLGSKGKRLRLPOIGSKNKL 1420
Db 842 SPSSLSANISSPKGSPSSSKGTSQSPSSKNSSPNSPRTLGSKGKRLRLPOIGSKNKP 901
QY 1421 SSSKENLDASKEGAGQICELADALSRLHVGSGQPELVTPQDHEVALANGFLYEHAAG 1480
Db 902 SSSKENLDASKEGAGQICELADALSRLHVGSGQPELVTPQDHEVALANGFLYEHAAG 961
QY 1481 N----GYSGNGQLGNSHREDSTDDRETRIKPIYVLAISCHSGILGGHYVYAKNPNC 1536
Db 962 NGCGGVSGNGQLGNSHREDSTDDRETRIKPIYVLAISCHSGILGGHYVYAKNPNC 1021
QY 1537 KWCYNSSCKELHDEIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSDDEDFESD 1596
Db 1022 KWCYNSSCKELHDEIDTDSAYILFYEQGIDYAOFLPKTDGKQWADTSSDDEDFESD 1081
QY 1597 YKTCVLIQ 1604
Db 1082 YEKYSMLQ 1089

RESULT 11

ABG21348
XX ABG21348 standard; Protein; 1125 AA.
XX AC ABG21348;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21339.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS85535.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 51707; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 1125 AA;
Query Match 51.8%; Score 4433.5; DB 22; Length 1125;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 842; Conservative 15; Mismatches 49; Indels 7; Gaps 4;
QY 701 EVRNKMSWPMSPIANSKIDRHKVPTEKGATGLNLGNTCFMNSSIOCVSNTPQLTQ 760
Db 220 EVRNKMSWPMSPIANSKIDRHKVPTEKGATGLNLGNTCFMNSSIOCVSNTPQLTQ 279
QY 761 YFISGRHLYELNRTNPIGMKGMAKCYGDLVQELWSGTOKNVAFLKLRWTIAKYAPFNG 820

280	YFISGRHLVELNRTNPIMKGHVAKYCGDLVQELWSGTQKSVAPLKRITIAKYAPKFG	339
821	FOQDSQELLAFLLDGLHEDLNRVHKPKYVELKSDSGRDPDWEAAEAWDNHLRNRSIVV	880
340	FOQDSQELLAFLLDGLHEDLNRVHKPKYVELKSDSGRDPDWEAAEAWDNHLRNRSIIV	399
881	DLPHGQLRSQVKCKTCGHSVRDPDNFNLSPLPMDSYMHELTIVIKLDGTTPVRYGLRL	940
400	DLPHGQLRSQVKCKTCGHSVRDPDNFNLSPLPMDSYMDELTIVIKLDGTTPVRYGLRL	459
941	NMDEKVTGLKKQLSDICGLNSBOILLAEVHGSNIKFPQDNQKRLSVSGFLCAFIPIVP	1000
460	NMDEKVTGLKKQLSDICGLNSBOILLAEVHGSNIKFPXDNQKRLSVSGFLCALEIPIVP	519
1001	VSPISASSPTQTDFFSSSPSTNEMFTLTITNGDLPRPFIPINGMENTVVPQTEKNTFNGMV	1060
520	VSPISASSPTQTDFFSPSTNEMFTLTITNGDLPRPFIPINGMENTVVPCTGDKNTFNGMV	579
1061	NGHMPSLPSPPTGYIIAVHRQMRTELIFLSQKQRPSLFGHPLIVPCTVTHTRKKOLDYD	1120
580	NGHMPSLPSPPTGYIIAVHRQMRTELIFLSQKQRPSLFGHPLIVPCTVTHTRKKOLDYD	639
1121	AVMIQVSRLASPLPPOEASNAHQDCDDSMGYQYFPTLRVVQKDGNSCAMCPWYRRCGCK	1180
640	AVMIQVSRLASPLPPOEASNAHQDCDDSMGYQYFPTLRVVQKDGNSCAMCPWYRRCGLX	699
1181	IDCGEDRAFTGNAYIAVDNDPTALHLRYQTSQERVVDEHESVQSRRAQAEPLNDSCLR	1240
700	IDCGEDRAFTGNAYIAVDNDPTALHLRYQTSQERVVDEHESVQSRRAQAEPLNDSCLR	759
1241	AFTSEBELGENEMYCSKCKTHCLATKGLDLMLPPLIIHLKRFQFVNGRWIKSQKIVK	1300
760	AFTSEBELGENEMYCSKCKTHCLATKGLDLMLPPLIIHLKRFQFVNDQMKSQKIVK	819
1301	FPRESFDPSAFLVPRPALCOHKPLTPQGDSEPRILAREVKKVDAAQSSAGIEDVLISK	1360
820	FPRESFDPSAFLVPRPALWQHPLTPQGDSEPRILAREVKKVDVSLAGEXDMLLSK	879
1361	SPSLSANIITSPKGPSRSRKSQSGTSCPSKSNSSPNSSPRTLGRSKGRRLRPLQIGSKNK	1419
880	SPSLSANIITSPKGPSRSRKSQSGTSCPSKSNSSPNKQPTGFWRSGRSGRLRPLQIGSKNK	939
1420	LSSSKENLDASKENGAGQICELADALSRGHVLGGSQPELVTPQDHEVALANGFLYEHEA	1478
940	PSSSKNLDASKENGAGQICELADALSRGHVRGVGQPELVTPQDHDVSLANGFLCEPEP	999
1479	CGN----GYNGQIGNHSEEDSTDQREDTRIKPIYNLYAISCHSGLGGHYVTVAKNP	1534
1000	CGNGCGDSSNGQLGNPSEARTDQREDTHXPIYNLXAISCHSETLSGGHYITTVAKNP	1059
1535	NCKWYCYNDSCKELHPDEITDSAYILFYEQQIDYAQFLPKTGDGKMADTSSMBDEDFE	1594
1060	NCKWYCYNDSCKELHDEITDSAYILFYEQQIDYAQFLPKTGDGKMADTSTDEBDE	1119
1595	SD-YKK 1599	
1120	SGLYKK 1125	

RESULT 12
ABG21344
ID ABG21344 standard; Protein; 1388 AA.
AC ABG21344;
XX
XX
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #21335.

WO200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Drmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
N-PSDB; AAS85531.
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
Claim 20; SEQ ID No 51703; 103pp; English.
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
Note: the sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
Sequence 1388 AA;

Query Match	50.5%;	Score 4328.5;	DB 22;	Length 1388;																																								
Best Local Similarity	90.0%;	Pred. No. 0;																																										
Matches 817;	Conservative 21;	Mismatches 29;	Indels 41;	Gaps 2																																								
QY	701	EVRKDMWP	EMSPFIANS	SKIDRHKVPT	KGATG	LSNLGNT	CFMNSS	TQCVSNT	QPTQ	760																																		
DB	518	EYHKDMWP	EMSPFTANS	SKIDRQKVPT	KGATG	LSNLGNT	CFMNSS	TQCVSNT	QPTQ	577																																		
QY	761	YFISGRHLY	ELNRNP	ICMGKHM	AKCVG	DDL	VQELWS	GTOKV	VAPI	KLRTIAKY	APRNG	820																																
DB	578	YFISGRHLY	ELNRNP	ICMGKHM	AKCVG	DDL	VQELNS	GTOK	SVAP	KLRTIAKY	APKPDG	637																																
QY	821	FOQDSQ	ELLAFL	LDGLH	DLNRV	HEKPY	VELKSD	GRPD	WEAA	EADNH	LRRNRS	IVV	880																															
DB	638	FOQDSQ	ELLAFL	LDGLH	DLNRV	HEKPY	VELKSD	GRPD	WEAA	EADNH	LRRNRS	IVV	697																															
QY	881	DLFHQL	RSQVK	CKTC	GHI	SVRP	DPN	FLSL	P	MD	SYMH	LEITV	IKL	DGTT	PVRY	GLL	940																											
DB	698	DLFHQL	RSQVK	CKTC	GHI	SVRP	DPN	FLSL	P	MD	SYMD	LEITV	IKL	DGTT	PVRY	GLL	757																											
QY	941	NWDEKY	TGLK	QK	LD	LCGL	NS	EQ	IL	LA	EVH	GS	NI	KNF	PP	QD	NQ	VLS	VS	GFL	CAFE	I	PVP	1000																				
DB	758	NWDEKY	TGLK	QK	LD	LCGL	NS	EQ	IL	LA	EVH	DS	NI	K										792																				
QY	1001	VSPIS	SASSP	TQ	DFSS	SP	STN	EM	PT	L	T	N	G	D	L	P	R	P	I	F	I	P	I	N	G	M	P	T	V	V	P	C	G	T	E	K	N	F	T	N	G	M	V	1061

RESULT 12	
ABG21344	
ID	ABG2:
XX	
AC	ABG2:

XZ	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.

Db 793 -----DSSSPSTNGMTLITNGDLPRFIPFGMPNTVVPCTGKNTFNGMV 840
Qy 1061 NGHMPSLPDSPTGYIIAVHRKMRMTLYFLSSQNRPSLFGMPLIVPCTVHTRKDLVD 1120
Db 841 NGHMPSLPDSPTGYIIAVHRKMRMTLYFLSQNRPSLFGMPLIVPCTVHTRKDLVD 900
Qy 1121 AVWLVSELASPLPOEASNAHQCDSDSMGYQYPTFLRVVQKDGNSCAWCPWTRFCRGCK 1180
Db 901 AVWLVSELASPLPOEASNAHQCDSDSMGYQYPTFLRVVQKDGNSCAWCPWTRFCRGCK 960
Qy 1181 IDCGEDRAFIAGNAYIADWDPTALHLRYQTSQERVDHESVEQSRRAQAEPLNLDSCLR 1240
Db 961 IDCGEDRAFIAGNAYIADWDPTALHLRYQTSQERVDHESVEQSRRAQAEPLNLDSCLR 1020
Qy 1241 AFTSEELGENEMYCCKKTHCLATKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVK 1300
Db 1021 AFTSEELGENEMYCCKKTHCLATKLDLWRLPPILIIHLKRFQVNDQWIKSQKIVR 1080
Qy 1301 FPRSPFSAFLVPRDPALCOHKPLTPQGBELSEPRILAREVKVDQAQSSAGEDVLLSK 1360
Db 1081 FLRESFSAFLVPRDPALCOHKPLTPQGBELSEPRILAREVKVDQAQSSAGEDVLLSK 1140
Qy 1361 SPSSLSANISSPKGSPSSSRKSGTSCPSSKNSSPNSSPRTLGRSKGRRLPQIGSKNKL 1420
Db 1141 SPSSLSANISSPKGSPSSSRKSGTSCPSSKNSSPNSSPRTLGRSKGRRLPQIGSKNKP 1200
Qy 1421 SSSKENLDASKENGAGQICELADALSRRHVLGGQPELVTPQDHEVALANGFLYEHAAG 1480
Db 1201 SSSKENLDASKENGAGQICELADALSRRHVLGGQPELVTPQDHEVALANGFLYEHAAG 1260
Qy 1481 N----GYSNGOLGNHSEDDTDQRETRIKPIYALYAIACHSGILGGGHVYVAKNPNC 1536
Db 1261 NGCGDGYNGOLGNHSEDDTDQRETRIKPIYALYAIACHSGILGGGHVYVAKNPNC 1320
Qy 1537 KWCYNDSSCKELHPDEIDTDSAVILYEQGIDYAOFLPKTDGKQWADTSSWDEPESD 1596
Db 1321 KWCYNDSSCKELHPDEIDTDSAVILYEQGIDYAOFLPKTDGKQWADTSSWDEPESD 1380
Qy 1597 YKCYCVLQ 1604
Db 1381 YEKYSMLQ 1388

RESULT 13
ID AAE19177 standard; Protein; 582 AA.

XX AAE19177;
XX AC
XX DT 21-MAY-2002 (first entry)
XX DE Human protease, PRTS-14 protein.
XX KW Human; protease; PRTS-14; enzyme; gastritis; cirrhosis; Crohn's disease;
KW gastroenteric disorder; autoimmune; inflammatory; cell proliferative;
KW cardiovascular; developmental; epithelial; neurological; reproductive;
KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;
KW anaemia; asthma; atherosclerosis; hypertension; myocardial infarction;
KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;
KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;
KW Pick's disease; infertility; vitiligo; drug screening; gene therapy.
XX OS Homo sapiens.
XX FN WO200208396-A2.
XX DT 31-JAN-2002.
XX PD 17-JUL-2001; 2001WO-US22397.
XX PF
XX PR 21-JUL-2000; 2000US-220063P.
XX PR 28-JUL-2000; 2000US-221680P.

PR 04-AUG-2000; 2000US-223544P.
PR 11-AUG-2000; 2000US-224717P.
PR 16-AUG-2000; 2000US-225988P.
PR 23-AUG-2000; 2000US-227568P.
XX (INCYTE GENOMICS INC.
XX PA
XX PI Deleage AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;
PI Tribouley CM, Das D, Kallik DA, Nguyen DB, Lee EA, Khan FA;
PI Yue H, Au-Young J, Griffin JA, Policky JL, Ramkumar J, Yang J;
PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;
PI Sanjanwala MS, Yao MG, Burford N, Walia NK, Lal P, Lee S, Todd S;
PI Lo TP, Tang YT, Elliott VS, Azimzai Y, Lu Y;
XX WPI; 2002-206082/26.
DR N-PSDB; AAD30581.
XX
PT New human protease polypeptide, useful in diagnosis, prevention and
PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,
PT cell proliferative, developmental, epithelial and neurological
PT disorders
XX
PS Claim 1; Page 152-153; 182pp; English.
XX
CC The invention relates to an isolated human protease polypeptide (PRTS).
CC PRTS protein and DNA are useful for diagnosing, treating and preventing
CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),
CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,
CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,
CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,
CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),
CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders
CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,
CC Parkinson's disease), and reproductive disorders (infertility). PRTS
CC protein is useful in a number of drug screening techniques and to
CC analyse the proteome of a tissue or cell type. PRTS DNA is useful for
CC creating knockin humanised animals or transgenic animals to model human
CC diseases, in somatic or germline gene therapy and in microarrays
CC utilising fluids or tissues from patients to detect altered PKIN
CC expression. The present sequence is human PRTS-14 protein.
XX
SQ Sequence 582 AA;

Query Match 36.6%; Score 3138; DB 23; Length 582;
Best Local Similarity 100.0%; Pred. No. 2.6e-270;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1023 MFTLTNGDLPRFIPFGMPNTVVPCTGKNTFNGMVNGHMPSLPDSPTGYIIAVHRK 1082
Db 1 MFTLTNGDLPRFIPFGMPNTVVPCTGKNTFNGMVNGHMPSLPDSPTGYIIAVHRK 60
QY 1083 MWRTLEYFLSSQNRPSLFGMPLIVPCTVHTRKDLVDVAVIQVSRLASPLPOEASNAH 1142
Db 61 MWRTLEYFLSSQNRPSLFGMPLIVPCTVHTRKDLVDVAVIQVSRLASPLPOEASNAH 120
QY 1143 QCDSDSMGYQYPTFLRVVQKDGNSCAWCPWTRFCRGCKIDCGEDRAFIAGNAYIADWDPT 1202
Db 121 QCDSDSMGYQYPTFLRVVQKDGNSCAWCPWTRFCRGCKIDCGEDRAFIAGNAYIADWDPT 180
QY 1203 ALHLRYQTSQERVDHESVEQSRRAQAEPLNLDSCLRRAFTSEELGENEMYCCKKTH 1262
Db 181 ALHLRYQTSQERVDHESVEQSRRAQAEPLNLDSCLRRAFTSEELGENEMYCCKKTH 240
QY 1263 CLATKKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVKFPRESFSAFLVPRDPALCOH 1322
Db 241 CLATKKLDLWRLPPILIIHLKRFQVNGRWIKSQKIVKFPRESFSAFLVPRDPALCOH 300
QY 1323 KPLTPQGBELSEPRILAREVKVDQAQSSAGEDVLLSKSPSSLSANISSPKGSPSSSRK 1382
Db 301 KPLTPQGBELSEPRILAREVKVDQAQSSAGEDVLLSKSPSSLSANISSPKGSPSSSRK 360
QY 1383 SGTCSPSSKNSSPNSSPRTLGRSKGRRLPQIGSKNKLSSKENLDASKENGAGQICEL 1442

QY 1002 -----SPISASSP----- 1009
Db 1091 DRAGKDELAPORNCYRMIYRWFCECEIRQESASVCFSTESLILSSSPENTFMH 1150
QY 1010 -----TOTDFSSSTNEMPTLTITNGDLPRIP-----NGMPT 1045
Db 1151 GAAQOKRVSSAKLHTESNTSSMGYTHSGENSWESSLTPEIPLADLEVPVSSRNGSGGE 1210
QY 1046 VPCGTEKFTNGMVGHM--PSL-----PDSPTG----- 1074
Db 1211 DCSYRTPNDSSGLTGTHTLGASLDVDEQAEGBNAEDHDQDQITTSQPTSSGVSRRS 1270
QY 1075 -----YIIVAVRKMVRELVELSSQKRPISLFGMLPLVPCVTRKDLDAVWIO 1125
Db 1271 SPPHAKGILVAVRKHTRDSYFSLVHKTRPSLFGVPLIPNSEGTHKDLCAVWLQ 1330
QY 1126 VSRASPLP-POEASNAHQDCCDSMGYQYPTFLRVQKDGNSCAWCPWYRCRCCKIDOG 1184
Db 1331 VSRLLSPLPATTEQANHAADCCDSGLGYDFPTLRAVADGLTCAICPWSFCRCCEIRCN 1390
QY 1185 EDRAFIG-----NA-----YI 1195
Db 1391 NDYVLOGALPINAASNTSTPKNAKPPSLNLEAKTPEYTAASLVTPTTKYFEDFTI 1450
QY 1196 AVMDPTALHLYQTSQBRVDEHESVQSRRAQAEPINLDSCLRAFTSBEELGENEMY 1255
Db 1451 AIDMDPTALHLYQSTLERLWVDHETIAISRRQVEPVDLNHCLRAFTSEKL--EOWIH 1508
QY 1256 CSKCTHCLATKLDLWRLPILILHLKRFVNGRWIKSQIYKFPRESFDPSPAPL--V 1313
Db 1509 CSHCKGKPKATKGLQIWKLPILILVHLKRFVNGRWIKSQIYKFPRESFDPSPAPL 1568
QY 1314 PRDPALCOHKPLTPQGDSESPRILAREVKVDAQSSAGEEDVLLSKSPSLNSNISSP 1373
Db 1569 PQETIL-RHKEL-----EL-----KNDAMTNAVNEV----- 1596
QY 1374 KSPSSSKSTSCPS--KNSSPNSPRTLGRSGRLRLPQIGSKNKLSSKENLDASK 1431
Db 1597 -----SELDEIDAPSKEVEELNQT-----GSKATASPPPTGNILRQSKTK----- 1639
QY 1432 ENGAGICELADALSRGHVLGSGPELVTPQDHEVALANGELYHEACGNGYNGQGNH 1491
Db 1640 -----NAVROKLSTSTK--TP-----IVDGEFEDYH-----QH 1668
QY 1492 SEEDSTDQREDTRIKTYNIYALISCHSGILGGHYVYAKNPCKWYCVNDSCKEL-H 1550
Db 1669 RLKPDVD-----QFDPYRLYAVVSHSGMLNGHYISYASNATGSGWYCVNDSCKELISQ 1722
QY 1551 PDEIDTSAYILFVEQOQIDYAOPLKPTDGHKMDTSSMD---EDFESDYKVC 1601
Db 1723 KPVIDPSAAYLLFYERKGLDYEPYLPNIEGRLTNTASVPLEVDETEGELKLC 1776

RESULT 15

AAU87244 standard; Protein, 539 AA.
AC AAU87244;
XX
DT 05-JUN-2002 (first entry)
DE Novel central nervous system protein #154.
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasia; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischemia; AIDS; angiodysplasia;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.

XX Homo sapiens.
XX WO200155318-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01332.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.

PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241788.
PR	20-OCT-2000;	2000US-0241789.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246509.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	06-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2000US-0254097.
PR	05-JAN-2001;	2001US-0253678.

PA	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Barash SC, Ruben SM;
PI	
XX	WPI; 2001-581633/65.
DR	N-PSDB; ABR43974.
XX	New isolated nucleic acid encoding a protein for diagnosing,
PT	preventing, treating or ameliorating medical conditions and used as
PT	food additives or preservatives -
XX	
XX	Claim 9; SEQ ID No 762; 837pp; English.
XX	The invention describes an isolated nucleic acid molecule (I) encoding a
CC	novel central nervous system protein. (I) and polypeptides (III) encoded
CC	by (I), are used to treat a medical conditions and in diagnosis of a
CC	pathological condition. Disorders which are diagnosed or treated include
CC	autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC	disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC	e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC	angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC	amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC	e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC	e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC	adenocarcinomas and irritable bowel syndrome, reproductive system
CC	disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC	and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC	leukaemia, disorders involving neovascularisation e.g. malignancies,
CC	respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC	acute kidney failure and blood related disorders e.g. myocardial
CC	infarction. The polypeptides can also be used to aid wound healing and
CC	epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC	maintain organs before transplantation, for supporting cell culture of
CC	primary tissues, to regenerate tissues and in chemotaxis. The
CC	polypeptides can also be used as a food additive or preservative to
CC	increase or decrease storage capabilities, fat content, lipid, protein,
Query Match	31.7%; Score 2711; DB 22; Length 539;
Best Local Similarity	93.3%; Pred. No. 3e-232;
Matches 512; Conservative	9; Mismatches 6; Indels 22; Gaps 2
QY	117 ERMHLVVGKVPDTLRKCFSGEKVNYEKEPNMLFLNKDAFTSRLLGGVVVLTDDSDS 176
Db	7 ERMHLVVGKVPDTLRKCFSGEKVNYEKEPNMLFLNKDAFTSRLLGGVVVLTDDSDS 66
QY	177 DPTFTFYQTLAGVTHLEESDIIDLEKRYWLKKAOSRTGRFDLETFGPLVSPRPSPSEGI 236
Db	67 DPTFTFYQTLAGVTHLEESDIIDLEKRYWLKKAOSRTGRFDLETFGPLVSPRPSPSEGI 126
QY	237 FNAFENRDNHIDFKEISCGLSACCRGPLEAQERKFCEKFPDVDRDGVLSRVELRDMMVAL 296
Db	127 FNAFENRDNHIDFKEISCGLSACCRGPLEAQERKFCEKFPDVDRDGVLSRVELRDMMVAL 186
QY	297 LEVWKONRTDOIPELHMDSLDSIVSGILNAHDTTKMGHLLTLEDYQIWSKNVLANEFINLL 356
Db	187 LEVWKONRTDOIPELHMDSLDSIVSGILNAHDTTKMGHLLTLEDYQIWSKNVLANEFINLL 246
QY	357 FOVCHIVILGLRPATPEEGQIIIRGWLERESRYGLQAHNWFIISMOWQKWKEYVKYKDAN 416
Db	247 FOVCHIVILGLRPATPEEGQIIIRGWLERESRYGLQAHNWFIISMOWQKWKEYVKYKDAN 306
QY	417 PVVIEPSSVLNGKGYSFCGTAAHPMEQVEDRIGSSLSYVNTTEKFSNDNSTASEASETAG 476
Db	307 PVVIEPSSVLNGKGYSFCGTAAHPMEQVEDRIGSSLSYVNTTEKFSNDNSTASEASETAG 366
QY	477 SGFLYSATPGADVCFAROHNTSDNNNCLLGANGNIILLNPQGAINDQPLVTQEPVK 536
Db	367 SGFLYSATPGADVCFAROHNTSDNNNCLLGANGNIILLNPQGAINDQPLVTQEPVK 426
QY	537 ATSLTLEGRLKRTPOLIHGRDYEMVPEPVWRALYHWYGANLAPRFVIKNSKTDPLEPE 596
Db	427 ATSLTLEGRLKRTPOLIHGRDYEMVPEPVWRALYHWYGANLAPRFVIKNSXTDIPLEPE 486

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:08:55 ; Search time 22 seconds
(without alignments)
3084.846 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKESRIGFLSYEALRRV.....DTSSMDEPESDYKVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/aaa/5A-COMB.pep:*
2: /cgn2_6/prodata/1/aaa/5B-COMB.pep:*
3: /cgn2_6/prodata/1/aaa/6A-COMB.pep:*
4: /cgn2_6/prodata/1/aaa/6B-COMB.pep:*
5: /cgn2_6/prodata/1/aaa/PCITUS-COMB.pep:*
6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2392.5	27.9	462	4	US-09-166-350-18
2	529	6.2	292	4	US-09-406-045-5
3	367	4.3	926	1	US-08-159-340A-2
4	347.5	4.1	193	1	US-08-159-340A-3
5	345	4.0	63	1	US-08-159-340A-5
6	334	3.9	67	1	US-08-159-340A-6
7	311	3.6	55	3	US-09-019-095A-32
8	286	3.3	54	1	US-08-159-340A-8
9	280	3.3	52	1	US-08-159-340A-7
10	217.5	2.5	1123	4	US-09-408-865-1
11	215	2.5	545	3	US-09-019-095A-38
12	201.5	2.4	509	3	US-09-019-095A-9
13	200	2.3	508	3	US-09-019-095A-8
14	200	2.3	521	3	US-09-019-095A-22
15	200	2.3	526	3	US-09-019-095A-2
16	184	2.1	418	3	US-09-019-095A-10
17	173	2.0	494	3	US-09-019-095A-24
18	160.5	1.9	220	4	US-09-399-913-24
19	160.5	1.9	220	4	US-09-298-731-24
20	156.5	1.8	191	3	US-08-655-352-5
21	156.5	1.8	191	3	US-08-655-352-6
22	156.5	1.8	191	4	US-09-258-016-5
23	156.5	1.8	191	4	US-09-258-016-6
24	156.5	1.8	191	4	US-09-257-825B-5
25	156.5	1.8	191	4	US-09-257-825B-6
26	156.5	1.8	220	4	US-09-399-913-26
27	156.5	1.8	220	4	US-09-298-731-26

28 155 1.8 291 4 US-09-167-206-22 Sequence 22, Appl
29 152.5 1.8 256 4 US-09-399-913-32 Sequence 32, Appl
30 152 1.8 118 3 US-08-905-223-399 Sequence 399, Appl
31 151.5 1.8 225 4 US-09-399-913-30 Sequence 30, Appl
32 151.5 1.8 225 4 US-09-398-731-30 Sequence 30, Appl
33 151.5 1.8 252 4 US-09-399-913-22 Sequence 22, Appl
34 151.5 1.8 252 4 US-09-399-913-28 Sequence 28, Appl
35 151.5 1.8 252 4 US-09-399-913-42 Sequence 42, Appl
36 151.5 1.8 252 4 US-09-298-731-22 Sequence 22, Appl
37 151.5 1.8 252 4 US-09-298-731-28 Sequence 28, Appl
38 151.5 1.8 257 4 US-09-399-913-16 Sequence 16, Appl
39 151.5 1.8 257 4 US-09-398-731-16 Sequence 16, Appl
40 151.5 1.8 270 4 US-09-399-913-18 Sequence 18, Appl
41 151.5 1.8 270 4 US-09-298-731-18 Sequence 18, Appl
42 151 1.8 193 3 US-08-655-352-4 Sequence 4, Appl
43 151 1.8 193 4 US-09-258-016-4 Sequence 4, Appl
44 151 1.8 193 4 US-09-257-825B-4 Sequence 4, Appl
45 150 1.8 174 1 US-08-328-322-17 Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-166-350-18
; Sequence 18, Application US/09166350A
; Patent No. 6440663

GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew

; APPLICANT: Chen, Yao

; APPLICANT: Stockert, Elisabeth

; APPLICANT: Old, Lloyd

; APPLICANT: Jager, Elke

; APPLICANT: Knuth, Alex

; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051

; CURRENT APPLICATION NUMBER: US/09/166,350A

; CURRENT FILING DATE: 1998-10-05

; EARLIER APPLICATION NUMBER: US 09/166,350

; EARLIER FILING DATE: 1998-10-05

; NUMBER OF SEQ ID NOS: 35

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 18

; LENGTH: 462

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-166-350-18

Query Match 27.9%; Score 2392.5; DB 4; Length 462;

Best Local Similarity 97.4%; Pred. No. 9.2e-201;

Matches 453; Conservative 1; Mismatches 6; Indels 5; Gaps 2;

QY 536 KATSLTLEGGELKRTQLHGRDYEMVPPVWRALYHWYGANLALPRPVIKSKTDIPEL 595
Db 1 KATSLTLEGGELKRTQLHGRDYEMVPPVWRALYHWYGANLALPRPVIKSKTDIPEL 60
QY 596 ELFPYLLFLRQOPATRTQOSNIWVWNGVPSNAPLKRVLAYTGCFSRMQTIKEIHEYL 655
Db 61 ELFPYLLFLRQOPATRTQOSNIWVWNGVPSNAPLKRVLAYTGCFSRMQTIKEIHEYL 120
QY 656 SQRLRKEEDMRUWLVNSENLYLTDDEHKLLEYLKIODEQHLVIEVRNKNKSWPEMSF 715
Db 121 SQRLRKEEDMRUWLVNSENLYLTDDEHKLLEYLKIODEQHLVIEVRNKNKSWPEMSF 180
QY 716 IANSSKIDRHKVPTEKATGLSLNGLNCTFWNSSIQCVSNTQPLTQYFISGRHLYELNRN 775
Db 181 IANSSKIDRHKVPTEKATGLSLNGLNCTFWNSSIQCVSNTQPLTQYFISGRHLYELNRN 240
QY 776 PIGMKGHMAKCYGDLVQELWSGTQKQVAPLKLRTWTAKYAPRNFNGQQDSOELLAFLLD 835
Db 241 PIGMKGHMAKCYGDLVQELWSGTQKQVAPLKLRTWTAKYAPRNFNGQQDSOELLAFLLD 300

QY 836 GLHEDLNVRHVKPYVELKDSGRDPDWEVAABWBNHLRRNSIYVDLPHQOLRSQVCKCT 895
DB 301 GLHEDLNVRHVKPYVELKDSGRDPDWEVAABWBNHLRRNSIYVDLPHQOLRSQVCKCT 360
QY 896 CGHISVRDPDNFSLPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSD 955
DB 361 CGHISVRDPDNFSLPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSD 420
QY 956 LGLNSEQILAEVHGNSIKNFPQDNQKVR--LSVSGFLCAFEIP 998
DB 421 LGLNSEQILAEVHGNSIKNFPQDNQKVR--LSVSGFLCAFEIP 998
RESULT 2
US-09-406-045-5
; Sequence 5, Application US/09406045
; Patent No. 6451994
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23413, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-50
; CURRENT APPLICATION NUMBER: US/09/406,045
; CURRENT FILING DATE: 1999-09-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ProDom consensus sequence
US-09-406-045-5

Query Match 6.2%; Score 529; DB 4; Length 292;
Best Local Similarity 23.1%; Pred. No. 1.1e-37;
Matches 137; Conservative 52; Mismatches 80; Indels 324; Gaps 10;
QY 732 GATGLNLGNTCFMSSQCVSNTPQLQYFISGRLHYEL--NRTNPIGMKGHAKYVGD 789
DB 9 GYTGLKNLGNTCFMSSQCVSNTPQLQYFISGRLHYEL--NRTNPIGMKGHAKYVGD 68
QY 790 LVQELWS-GTQKNVAPKLRTIAYAPFN-GFOQDSQELLAFLLDGLHEDLNVRHVK 847
DB 69 LVHQMNSNSKSVAPQLTIVGVFSQFSGYQQQSQFLKFLQDDAHEDFNLMEX 128
QY 848 PYVE--LKDSGKPD--MEVAEAENDLRRNSIYVDLPHQOLRSQVCKCTGCHIS-VR 902
DB 129 PYVEEQVDSNEKSTALNVVSEAEWENHKRNDISIIFQCFKSTIKCPSCHEITSETT 188
QY 903 FDFNFSLPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSDLCGLNSE 962
DB 189 FDFNFSLPLPMDSYMHLEITVTKLDTTPVRYGLRLNMDKTYGLKKQLSDLCGLNSE 203
QY 963 QILAEVHGNSIKNFPQDNQKVRLSVSGFLCAFEIPVFPVSPISASSP*QTDFFSSSPSTNE 1022
DB 204 ----- 203
QY 1023 MFTLTNGDLPRPIFIPNGMNTVVPCTGKFTNGMNGMPSLPDPSFTGYIIAIVHRK 1082
DB 204 ----- 203
QY 1083 MMRTELYFLSSQKNSPLFGMPLIVPCTVHTRKDLDAVWIVQVSRSLASPLPQASNHA 1142
DB 204 -----ADNHQ 208
QY 1143 QDCDDSMGYQFPTLRVQKQNSCAWCPWYFCRGCKIDCGEDRAFIAGYAVIADWDPT 1202
DB 209 ----- 208
QY 1203 ALHLRYCTQERVVDESHVESQSRRAQAPINLDSCLRAFTSEELGENEMYCKSKKTH 1262
DB 209 -----NLQDCLESFTKKELEGDNKWCYCFKCKK 237

QY 1263 CLATKKLDLRLPPLIILHLKRFQVNGRWIKSOKI---VKFPRESFDPASFL 1312
DB 238 QEATKKLDLWLKPPVIVILHLKRFYSI-DQWGRDRKINTTVEFFLEDLMSPVY 289
RESULT 3
US-08-159-340A-2
; Sequence 2, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; APPLICANT: Papa, Feroz
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-2

Query Match 4.3%; Score 367; DB 1; Length 926;
Best Local Similarity 21.8%; Pred. No. 1.5e-22;
Matches 159; Conservative 100; Mismatches 225; Indels 242; Gaps 30;
QY 242 ENRDNHDIPKEISCGLSACRCGRPLAERQKFCVDPDVRDGVLSRVELRDMVALLEYWK 301
DB 219 ENKSHDITKNIIC-----LEPISFKMSYSDHD-----LEKSLI----- 253
QY 302 DNRTDDIPELHMDLS-DIVEGILNAHDTTKMGHLATLEDYQIWSKXNLANEFNLFFQVC 360
DB 254 ---TSNSEIKMFQSRNLFKFIILYDA-----NEYNV-KQOSVLLDILVNSHFE-- 299
QY 361 HVLGLRPAATPE-----BEGQIRGWLERSRYGLQAGHWFIIISMQWQWQKEYVY 413
DB 300 -----KPISDDTKFIIESG--PPGWL--KSNYGRQVSSS-----F 332
QY 414 DANPVVIEPSSVLNGKYGFGTAHHPMEQVEORIGSS-----LSYVNTTEE 459
DB 333 PSNNIKDDSVIYNGN--TSGLSLQHLPKMSIRHSMDSDMKMLVAPTPLNHLQOQQ 390
QY 460 KFDENISTASEAETAGSGFLYSATPGADVCFARQNTSDNNQCLLGANGNILLHLNPO 519
DB 391 QQSDNDHVLKRSS-----SFKKLFENYTPSPKNSNLSYISLSISSSPS 437
QY 520 KPGALDNPQVLTQEPVKATSLTEGGRLKRTPOLIHDRDYEMVPEPVRALYHWYGANLA 579

Db 438 P-----LPLHSPDPVKNSFRI---NYBPETHL-----W-----463
QY 580 LRPVINKNSKTDPLELEPRYLFLRQOPATFQOSNIWVMGNVPSPNAPLKRVLAYT 639
Db 464 -----KNSBTD-----FVINOQELNHSFAHIAFIPATKATSP-----497
QY 640 GCFSRMOT--IKEIHYLSORLAKEEDMRMLYNSENYLTLLDDDHKLKYLKIQDECH 697
Db 498 ---SRTATPKLQRPPTQISNMNL-----MNSNGHSSATSTIQ-----531
QY 698 LVTEVNKMSWEEMSFANSKSIDRHKV--PTEKG-----ATGLSNLGNTCFNMSSIQ 750
Db 532 -----PSCUS--LSNNDLSDHTDVTPTSSHNYDLDFAVGLENLGNSCYMNCIIQ 578
QY 751 CVSNTQPLTOYFTSGRHLVELNRTNPIGMKGMAKCYGDLVQELW-----SGTOK-NVAPL 805
Db 579 CILGTHLTIQIFLDDSYAKHININSLKSGILAKYFAELVHMYKEQVDGSKKISISPI 638
QY 806 KLRWTIAYAPRNGFQQODSOELLAFLLDGLHEDLNVRVHEKPYVELKDSGRDPWEVAA 865
Db 639 KFKLACGSVNSLFTASQODCOFCOFLDGLHEDLNQCGSNP-----PLKELSQ 688
QY 866 EA-----WDNHLRRNRSIVVDLFGHQLRSQVCKCTCGHISVRFDPENFL 909
Db 689 EABAREKLSLRASSIEWERFLTDFSVIVDLFQOYASRLKCKVCSHTTITTYQFTVL 748
QY 910 SLPLP 914
Db 749 SIIPI 753
RESULT 4
US-08-159-340A-3
; Sequence 3, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-3

Query Match 4.1%; Score 347.5; DB 1; Length 193;
Best Local Similarity 36.5%; Pred. No. 4.2e-22;
Matches 74; Conservative 31; Mismatches 67; Indels 31; Gaps 4;
QY 733 ATGLSNLGNTCFNMSSIQCVSNTQPLTOYFTISGRHLVELNRTNPIGMKGMAKCYGDLVQ 792
Db 1 AVGLENLGNSCYMNCIIQCIILGTHLTIQIFLDDSYAKHININSLKSGILAKYFAELVH 60
QY 793 ELW-----SGTOK-NVAPLKLRTWTIAYAPRNGFQQODSOELLAFLLDGLHEDLNVRVHEK 847
Db 61 MYKEQVDGSKKISISPIKFKLACGSVNSLFTASQODCOFCOFLDGLHEDLNQCGSN 120
QY 848 PYVELKDSGRDPWEVAAEA-----WDNHLRRNRSIVVDLFGHQLRSQV 891
Db 121 P-----PLKELSQEABAREKLSLRASSIEWERFLTDFSVIVDLFQOYASRL 170
QY 892 KCKTCTGHISVRFDPENFLSLPLP 914
Db 171 KCKVCSHTTITTYQFTVLSIPI 193
RESULT 5
US-08-159-340A-5
; Sequence 5, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-5
Query Match 4.0%; Score 345; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 8.7e-23;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 733 ATGLSNLGNTCFNMSSIQCVSNTQPLTOYFTISGRHLVELNRTNPIGMKGMAKCYGDLVQ 792
Db 1 ATGLSNLGNTCFNMSSIQCVSNTQPLTOYFTISGRHLVELNRTNPIGMKGMAKCYGDLVQ 60
QY 793 ELW 795
|||

CURRENT APPLICATION NUMBER: US/09/019,095A
CURRENT FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 38
LENGTH: 545
TYPE: PRT
ORGANISM: murine
US-09-019-095A-38

Query Match 2.5%; Score 215; DB 3; Length 545;
Best Local Similarity 28.5%; Pred. No. 1.2e-09;
Matches 65; Conservative 32; Mismatches 79; Indels 52; Gaps 8;
QY 693 QDEQHLVIEVRKND---MSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFNMSSI 749
DB 24 QDEAQQVVELTANDKESLSW-----ECQGPQC-GLQNTGNSCYLNAAL 66
QY 750 QCVSNTQPLTQYFISGRHLYELNRTNPIGMKGMKCYGDLVQELWSGTQKNVAPLKL 809
DB 67 QCLTHPTPLADYMLSOE--YSQCCSPGCK--MCAMEALVTQSLHSHSGDVMKPSQIL 122
QY 810 TIATYAPRNFNGFOQDSQELALFLDLGLHEDLNRVHEKPYVELKSDGCRPDVEVAEAWD 869
DB 123 TSA-----FHKQOQDAHEFLMTLTHSHSCLOVHRQSEPTSDS----- 163
QY 870 NHLNRNSIVVDLPHGQLRSQVKCKTCGHSVRFPDFNLSPLPMD 917
DB 164 -----SPIHDFGLWRWSQIKCLHCOQTSPTYD--RFLDPLDISS 202

RESULT 12

US-09-019-095A-9
Sequence 9, Application US/09019095A
Patent No. 6287858
GENERAL INFORMATION:
APPLICANT: D'Andrea, Alan D.
APPLICANT: Zhu, Yuan
TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
TITLE OF INVENTION: Cell Growth
FILE REFERENCE: DFCI-435p2A2
CURRENT APPLICATION NUMBER: US/09/019,095A
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 9
LENGTH: 509
TYPE: PRT
ORGANISM: Murine
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(509)
NAME/KEY: VARIANT
LOCATION: (1)...(509)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-019-095A-9

Query Match 2.4%; Score 201.5; DB 3; Length 509;
Best Local Similarity 31.6%; Pred. No. 1.6e-08;
Matches 60; Conservative 30; Mismatches 53; Indels 47; Gaps 10;

QY 735 GLSLNLTGTCFNMSSIQCVSNTQPLTQYFISGRHLYELNRT--NPIGMK-GHMAKCYGDLV 791
DB 41 GLQNTGNSCYLNAALQCLTHPTPLADYMLSOE-----SQCCSPGCKMCAMEEC---VT 93
QY 792 QELWSGTQKNVAPLKLRTIAK---YAPRNFNGFOQDSQELALFLDLGLHEDLNRVHEK 847
DB 94 QSL-----XLSLXGDVMKPSQILTSFAFHKQOQDAHEFLMTLTHSHSCLOVH-- 142
QY 848 PYVELKSDGCRPDVEVAEAWDNHLNRNSIVVDLPHGQLRSQVKCKTCGHSVRFPDFN 907
DB 143 -----RQSDPTP-----QDTSPIHDFGWRWSQIKCLXAG--TSHTTDP-- 180
QY 908 FLSPLPMD 917
DB 181 FLDVPLDISS 190

RESULT 13

US-09-019-095A-8
Sequence 8, Application US/09019095A
Patent No. 6287858
GENERAL INFORMATION:
APPLICANT: D'Andrea, Alan D.
APPLICANT: Zhu, Yuan
TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
TITLE OF INVENTION: Cell Growth
FILE REFERENCE: DFCI-435p2A2
CURRENT APPLICATION NUMBER: US/09/019,095A
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 508
TYPE: PRT
ORGANISM: Murine
US-09-019-095A-8

Query Match 2.3%; Score 200; DB 3; Length 508;
Best Local Similarity 27.8%; Pred. No. 2.1e-08;
Matches 64; Conservative 34; Mismatches 76; Indels 56; Gaps 9;

QY 693 QDEQHLVIEVR---NKDMSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFNMSSI 749
DB 13 QDEAQQVVELTANDKESLSW-----ESQGPQC-GLQNTGNSCYLNAAL 55
QY 750 QCVSNTQPLTQYFISGRHLYELNRT--NPIGMKGMKCYGDLVQELWSGTQKNVAPLKL 807
DB 56 QCLTHPTPLADYMLSOE-----SQCCSPGCK--LCAMEALVTQSLHSHSGDVMKPSH 109
QY 808 RWTIATYAPRNFNGFOQDSQELALFLDLGLHEDLNRVHEKPYVELKSDGCRPDVEVAE 867
DB 110 ILTSA-----FHKQOQDAHEFLMTLTHSHSCLOVHRQSEPTSDS----- 152
QY 868 WDNHLNRNSIVVDLPHGQLRSQVKCKTCGHSVRFPDFNLSPLPMD 917
DB 153 -----SPIHDFGWRWSQIKCLHCOQTSPTYD--RFLDPLDISS 191

RESULT 14

US-09-019-095A-22
Sequence 22, Application US/09019095A
Patent No. 6287858
GENERAL INFORMATION:
APPLICANT: D'Andrea, Alan D.
APPLICANT: Zhu, Yuan
TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
TITLE OF INVENTION: Cell Growth

FILE REFERENCE: DFCI-435p2a2
CURRENT APPLICATION NUMBER: US/09/019,095A
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 22
LENGTH: 521
TYPE: PRT
ORGANISM: Murine
US-09-019-095A-22

Query Match 2.3%; Score 200; DB 3; Length 521;
Best Local Similarity 27.8%; Pred. No. 2.2e-08;
Matches 64; Conservative 34; Mismatches 76; Indels 56; Gaps 9;
QY 693 QDEQHLVIEVR---NKDMSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFMNSSI 749
DB 19 QDEAQVVEELTVNGKHSLSW-----ESPQPGC-GLQNTGNSCYLNAAL 61
QY 750 QCVSNTQPLTQYFISGRHLYELNRT--NPIGMKGHMAKCYGDLVQELWSGTQKNVAPLKL 807
DB 62 QCLTHTPPLADYMLSQEH-----SQTCCSPGCK--LCAMEALVTQSLHSHSGDVMPKPSH 115
QY 808 RWTIAKYAPRFNGFQQODSOELLAFLLDGLHEDLNVRVHEKPYVELKDSGRPDWEVAAEA 867
DB 116 ILTSA-----FKHQQEDAHEFLMFTLETWHESCLQVHRQSKPTSEDS----- 158
QY 868 WDNHLRRNSIVVDLPHGQLRSQVKCTCGHISVRFPDPFNFLSLPLPMDS 917
DB 159 -----SPIHDIRGWMRSQIKCLLCQGTSDTYD--RFLDIPLDISS 197

RESULT 15

US-09-019-095A-2
Sequence 2, Application US/09019095A
Patent No. 6287858
GENERAL INFORMATION:
APPLICANT: D'Andrea, Alan D.
APPLICANT: Zhu, Yuan
TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
FILE REFERENCE: DFCI-435p2a2
CURRENT APPLICATION NUMBER: US/09/019,095A
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 526
TYPE: PRT
ORGANISM: Murine
US-09-019-095A-2

Query Match 2.3%; Score 200; DB 3; Length 526;
Best Local Similarity 27.8%; Pred. No. 2.3e-08;
Matches 64; Conservative 34; Mismatches 76; Indels 56; Gaps 9;
QY 693 QDEQHLVIEVR---NKDMSWPEMSFIANSKIDRHKVPTEKGATGLSLNLTGTCFMNSSI 749
DB 24 QDEAQVVEELTVNGKHSLSW-----ESPQPGC-GLQNTGNSCYLNAAL 66
QY 750 QCVSNTQPLTQYFISGRHLYELNRT--NPIGMKGHMAKCYGDLVQELWSGTQKNVAPLKL 807

DB 67 QCLTHTPPLADYMLSQEH-----SQTCCSPGCK--LCAMEALVTQSLHSHSGDVMPKPSH 120
QY 808 RWTIAKYAPRFNGFQQODSOELLAFLLDGLHEDLNVRVHEKPYVELKDSGRPDWEVAAEA 867
DB 121 ILTSA-----FKHQQEDAHEFLMFTLETWHESCLQVHRQSKPTSEDS----- 163
QY 868 WDNHLRRNSIVVDLPHGQLRSQVKCTCGHISVRFPDPFNFLSLPLPMDS 917
DB 164 -----SPIHDIRGWMRSQIKCLLCQGTSDTYD--RFLDIPLDISS 202

Search completed: February 11, 2004, 15:32:25
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:30:16 ; Search time 45 Seconds

(without alignments)
7463.315 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKESRIGFLSYVEALRRV.....DTSSMDEDFSDYKVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTU5_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8563	100.0	1604	9	US-09-888-615-73
2	6826	79.7	1317	12	US-10-071-275-2
3	4428	51.7	847	12	US-10-094-749-3045
4	3138	36.6	582	12	US-10-274-639-14
5	2711	31.7	539	12	US-09-764-875-762
6	1750	20.4	401	12	US-10-158-057-222
7	1406	16.4	786	15	US-10-071-838-17
8	1124	13.1	1041	12	US-10-369-493-5738
9	994	11.6	827	12	US-10-210-130-84
10	890	10.4	173	11	US-09-764-881-153
11	890	10.4	173	12	US-10-243-747-153
12	890	10.4	173	12	US-10-158-057-300
13	890	10.4	173	12	US-09-764-875-1064
14	787.5	9.2	978	12	US-10-369-493-22797
15	756.5	8.8	900	12	US-10-369-493-6013

16	738.5	8.6	1254	12	US-10-369-493-22131	Sequence 22131, A
17	675	7.9	1095	12	US-10-369-493-5583	Sequence 5583, Ap
18	665	7.8	924	12	US-10-369-493-22682	Sequence 22682, A
19	665	7.8	1318	12	US-10-291-172-338	Sequence 338, App
20	646.5	7.5	929	12	US-10-369-493-3912	Sequence 3912, A
21	530.5	6.2	680	12	US-10-369-493-12503	Sequence 12503, A
22	529	6.2	292	14	US-10-105-992-5	Sequence 5, Appli
23	524	6.1	1123	9	US-09-888-615-77	Sequence 77, Appli
24	498	5.8	441	15	US-10-153-668-102	Sequence 102, App
25	498	5.8	634	15	US-10-153-668-98	Sequence 98, Appl
26	498	5.8	634	15	US-10-153-668-100	Sequence 100, App
27	498	5.8	1118	15	US-10-153-668-104	Sequence 104, App
28	488	5.7	1080	12	US-10-205-219-62	Sequence 62, Appl
29	408	4.8	605	12	US-10-094-749-2994	Sequence 2994, Ap
30	371	4.3	926	12	US-10-369-493-1623	Sequence 1623, Ap
31	330	3.9	805	12	US-10-369-493-1769	Sequence 1769, Ap
32	321	3.7	565	10	US-09-766-954A-2	Sequence 2, Appli
33	318.5	3.7	489	12	US-09-833-245-211	Sequence 211, App
34	314	3.7	463	12	US-09-833-245-212	Sequence 212, App
35	288	3.4	435	12	US-10-264-237-2761	Sequence 2761, Ap
36	280	3.3	814	9	US-09-808-568-5	Sequence 5, Appli
37	280	3.3	814	12	US-10-098-108-7	Sequence 7, Appli
38	280	3.3	814	15	US-10-156-239-52	Sequence 52, Appl
39	273.5	3.2	812	15	US-10-097-340-330	Sequence 330, App
40	271	3.2	593	9	US-09-872-153-18	Sequence 18, Appl
41	257.5	3.0	207	12	US-10-264-049-4055	Sequence 4055, Ap
42	257.5	3.0	953	9	US-09-888-615-66	Sequence 66, Appl
43	249.5	2.9	1071	12	US-10-369-493-22077	Sequence 22077, A
44	249	2.9	656	12	US-10-274-639-20	Sequence 20, Appl
45	237.5	2.8	783	9	US-09-888-615-67	Sequence 67, Appl

ALIGNMENTS

RESULT 1

US-09-888-615-73
; Sequence 73, Application US/09888615
; Patent No. US20020044858A
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARVDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 73
; LENGTH: 1604
; TYPE: PRT
; ORGANISM: Homo sapiens
US-888-615-73

Query Match	100.0%;	Score 8563;	DB 9;	Length 1604;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1604;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGAKESRIGFLSYVEALRRVTDVELKELKDAFKRTGCLSYMGHCHCFIREVLGDGVPVKY 60		
Db	1	MGAKESRIGFLSYVEALRRVTDVELKELKDAFKRTGCLSYMGHCHCFIREVLGDGVPVKY 60		
Qy	61	AEVITYCSFGTSGKLHNNLIIVGLVLLTRCKDEKAKYIFSLFSSSGNVIREEVERML 120		
Db	61	AEVITYCSFGTSGKLHNNLIIVGLVLLTRCKDEKAKYIFSLFSSSGNVIREEVERML 120		
Qy	121	HVVDGKVPDTRLCFSGEKGKVEKFNWFLNKDAFTFSRWLLSGGVYVTLTDDSDTPT 180		

121 HVVDGKVPDLKCRSEGEKQYEFKRWLFLNKDAFTSRWLLSGGVYVTLTDSPT 180
 181 FYQTLAGVTHLESDDIIDEKRYWLLKAQSRGTREDFLTFGLVSPPIRPSLSGLFNAF 240
 181 FYQTLAGVTHLESDDIIDEKRYWLLKAQSRGTREDFLTFGLVSPPIRPSLSGLFNAF 240
 241 DENRDNHIDFKISGLSACCGPLAEROKFCFKVFDVDRDGVLSRVELRWVVALLEV 300
 241 DENRDNHIDFKISGLSACCGPLAEROKFCFKVFDVDRDGVLSRVELRWVVALLEV 300
 301 KNRRTDDIPELHMDLSDIIEGINAHDITKMGHLTLEDYQIWSVKNVLANFLNLLFOVC 360
 301 KNRRTDDIPELHMDLSDIIEGINAHDITKMGHLTLEDYQIWSVKNVLANFLNLLFOVC 360
 361 HVLGLRPATPEEGQIRGMLERSRYGLQAGHWNFIISQWQWQWKEYKYDANPVI 420
 361 HVLGLRPATPEEGQIRGMLERSRYGLQAGHWNFIISQWQWQWKEYKYDANPVI 420
 421 EPSVVLNGGKYSGTAAHPMEQVEDRIGSSLSVYNTTEKFSDNISTASEASTAGSGL 480
 421 EPSVVLNGGKYSGTAAHPMEQVEDRIGSSLSVYNTTEKFSDNISTASEASTAGSGL 480
 481 YSATPGADVCFARQHTSDNNQCLLGANGNTLLHNPQKGAIDNPQVYKATSL 540
 481 YSATPGADVCFARQHTSDNNQCLLGANGNTLLHNPQKGAIDNPQVYKATSL 540
 541 TLEGGRLKRTPOLIHDRDYMVPEPVWVRLYHWYGANLALPRVINKSKTIDPELELPPR 600
 541 TLEGGRLKRTPOLIHDRDYMVPEPVWVRLYHWYGANLALPRVINKSKTIDPELELPPR 600
 601 YLLFLRQPATRTOQSNVWNGVSPNAPLKRVLAYTGCFSRMTQIHEIYLSQRLR 660
 601 YLLFLRQPATRTOQSNVWNGVSPNAPLKRVLAYTGCFSRMTQIHEIYLSQRLR 660
 661 IKEEDMRLWYNSYNTLLDDEHKLKYLKIQDOHLYEVNKMDSKPMESFIANS 720
 661 IKEEDMRLWYNSYNTLLDDEHKLKYLKIQDOHLYEVNKMDSKPMESFIANS 720
 721 KIDRHVPTKEGATGSLNLCNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTPIGMK 780
 721 KIDRHVPTKEGATGSLNLCNTCFMNSSIQCVSNTQPLTQYFISGRHLYELNRTPIGMK 780
 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTIAYAPRNGFOQDSQELLAFLLDGLHED 840
 781 GHMAKCYGDLVQELWSGTQKNVAPLKLRTIAYAPRNGFOQDSQELLAFLLDGLHED 840
 841 LNRVHEKPYVELKDSGRPDWEVAAEAWNDHLRNRISIVVDLPHGLRSQVCKCTCGHIS 900
 841 LNRVHEKPYVELKDSGRPDWEVAAEAWNDHLRNRISIVVDLPHGLRSQVCKCTCGHIS 900
 901 VRDPDNFVLSPLPMDSYMHLEITVTKLGGTTPVRYGLRLNDEKVTGLKQSLDGLGN 960
 901 VRDPDNFVLSPLPMDSYMHLEITVTKLGGTTPVRYGLRLNDEKVTGLKQSLDGLGN 960
 961 SEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEIPVPVSPISASSPTQDRSSSPT 1020
 961 SEQILLAEVHGSNIKNFPDQNKVRLSVSGFLCAFEIPVPVSPISASSPTQDRSSSPT 1020
 1021 NEMFTLTNGDLPRPIFIPNGMNTVPCGTEKNFTNGMVGHPMLSPDPTGTIIAVH 1080
 1021 NEMFTLTNGDLPRPIFIPNGMNTVPCGTEKNFTNGMVGHPMLSPDPTGTIIAVH 1080
 1081 RKMRTYFLPSQKRPISLFGMPLVPCVTHTKKDLYDAVWIOVSRLASPLPQEAAS 1140
 1081 RKMRTYFLPSQKRPISLFGMPLVPCVTHTKKDLYDAVWIOVSRLASPLPQEAAS 1140
 1141 HAQCDSDSMGYQYFPTLRVYQKDGNSCAMCPWYRRCGCKIDCGEDRAFINAYIAVDWD 1200
 1141 HAQCDSDSMGYQYFPTLRVYQKDGNSCAMCPWYRRCGCKIDCGEDRAFINAYIAVDWD 1200
 1201 PTALHLRYQTSQREVDEHESVQSRRAQAPINLDSCLRAFTSBEELGENEMYCSCK 1260

1201 PTALHLRYQTSQREVDEHESVQSRRAQAPINLDSCLRAFTSBEELGENEMYCSCK 1260
 1261 THCLATKLDLWRLPILLIHLKRFQFVNGRWIKSOKIVKFPRESPPSAFLVPRDPALC 1320
 1261 THCLATKLDLWRLPILLIHLKRFQFVNGRWIKSOKIVKFPRESPPSAFLVPRDPALC 1320
 1321 QHKLPLTPQGDDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSSISANIISSPKGSPSS 1380
 1321 QHKLPLTPQGDDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSSISANIISSPKGSPSS 1380
 1381 RKSGTSCPSKNSPNSPRTLGRSKRLRLPOIGSKNKLSSKENLDAKNGAGQICE 1440
 1381 RKSGTSCPSKNSPNSPRTLGRSKRLRLPOIGSKNKLSSKENLDAKNGAGQICE 1440
 1441 LADALSRGHVILGSGQBELVTPQDHEVALANGFLYEACNGVSNQCLGNHSEEDSTDQ 1500
 1441 LADALSRGHVILGSGQBELVTPQDHEVALANGFLYEACNGVSNQCLGNHSEEDSTDQ 1500
 1501 REDTRIKPIYNLVYAISCHSGILGGGHVYTYAKNPCKWICYNDSSCKELHPDEIDTDSAY 1560
 1501 REDTRIKPIYNLVYAISCHSGILGGGHVYTYAKNPCKWICYNDSSCKELHPDEIDTDSAY 1560
 1561 ILFYEQGGIDYAOFLPKTDGKWADTSSMDEDESDYKKYCVLQ 1604
 1561 ILFYEQGGIDYAOFLPKTDGKWADTSSMDEDESDYKKYCVLQ 1604

RESULT 2

US-10-071-275-2
 ; Sequence 2, Application US/10071275
 ; Publication No. US20030138934A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meyers, Rachel E.
 ; TITLE OF INVENTION: 80091, A NOVEL HUMAN UBIQUITIN
 ; TITLE OF INVENTION: CARBOXY-TERMINAL HYDROLASE FAMILY MEMBER AND USES THEREOF
 ; FILE REFERENCE: 10448-139001
 ; CURRENT APPLICATION NUMBER: US/10/071,275
 ; PRIOR FILING DATE: 2002-02-07
 ; PRIOR FILING DATE: 2001-02-07
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1317
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-071-275-2

Query Match 79.7%; Score 6826; DB 12; Length 1317;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1274; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 330 KMGHLTLEDYQIWSVKNVLANFLNLLFOVCHIVLGLRPATPEEGQIRGMLERSRYG 389
 DB 43 QMGHLTLEDYQIWSVKNVLANFLNLLFOVCHIVLGLRPATPEEGQIRGMLERSRYG 102
 QY 390 LQAGHWNFIISQWQWQWKEYKYDANPVVIBESSVLNGKYSFGTAAHPMEQVEDRIGS 449
 DB 103 LQAGHWNFIISQWQWQWKEYKYDANPVVIBESSVLNGKYSFGTAAHPMEQVEDRIGS 162
 QY 450 SLSVYNTTEKFSDNISTASEASTAGSGFLYSATPGADVCFARQHTSDNNQCLLGAN 509
 DB 163 SLSVYNTTEKFSDNISTASEASTAGSGFLYSATPGADVCFARQHTSDNNQCLLGAN 222
 QY 510 GNILLHLNPKGAINDNPLOVTOEPVKATSLTLEGRRLKRTPOLIHDRDYMVPEPVWRA 569
 DB 223 GNILLHLNPKGAINDNPLOVTOEPVKATSLTLEGRRLKRTPOLIHDRDYMVPEPVWRA 282
 QY 570 LYHWYGANLALPRVINKSKTIDPELELFPYLLFLRQKQPATRTOQSNVWNGVSPN 629
 DB 283 LYHWYGANLALPRVINKSKTIDPELELFPYLLFLRQKQPATRTOQSNVWNGVSPN 342
 QY 630 APLKRVLAYTGCFSRMTQIHEIYLSQRLRIKEEDMRLWYNSYNTLLDDEHKLK 689

Db 343 APLKRVLAYTCFERNQTIKEIHLYLQRIKEEDNRLWLYNSENTLTLLDDDEHKEY 402
Qy 690 LKIQDEHLVTEVNKDKMSPFEMSFANGSKIDRHKVTEKAGTGLSNLNTCFMNSSI 749
Db 403 LKIQDEHLVTEVNKDKMSPFEMSFANGSKIDRHKVTEKAGTGLSNLNTCFMNSSI 462
Qy 750 QCVNTQPLTOYFISGRHLVELNTPNIGMKHMAKCYGDLVQELKSGTQKNVAPLKLW 809
Db 463 QCVNTQPLTOYFISGRHLVELNTPNIGMKHMAKCYGDLVQELKSGTQKNVAPLKLW 522
Qy 810 TIAYAPRFNGFOQDQOELLAFLLDGLHEDLNKVPVELKSDGRPDWEAAEAWD 869
Db 523 TIAYAPRFNGFOQDQOELLAFLLDGLHEDLNKVPVELKSDGRPDWEAAEAWD 582
Qy 870 NHLRNEISIVVDLPHGOLRSOVCKTGGHISVREDPENFLSLPLPMDSYNHLBITVILKD 929
Db 583 NHLRNEISIVVDLPHGOLRSOVCKTGGHISVREDPENFLSLPLPMDSYNHLBITVILKD 642
Qy 930 GTTPVRYGLRLNMDKTYGLKKQLSDLCGLNSQIILLAEVHGNIKNFPDQNKVRLSVS 989
Db 643 GTTPVRYGLRLNMDKTYGLKKQLSDLCGLNSQIILLAEVHGNIKNFPDQNKVRLSVS 702
Qy 990 GFLCAPIPVPSISASPTQDFFSSPSTNEMFTLTNGDLPRPIFIPNGMNTVWPC 1049
Db 703 GFLCAPIPVPSISASPTQDFFSSPSTNEMFTLTNGDLPRPIFIPNGMNTVWPC 762
Qy 1050 GTEKNFTNGMNGHMPSLPSPFTGYIIAHRKMWRTLYFLSKOKRPSLFCWPLIVPC 1109
Db 763 GTEKNFTNGMNGHMPSLPSPFTGYIIAHRKMWRTLYFLSKOKRPSLFCWPLIVPC 822
Qy 1110 TVHTRKDLDAVMIQVSRLASPLPPOEASNHQDCDDSMGYQYPTFLRVVQDGNSCAW 1169
Db 823 TVHTRKDLDAVMIQVSRLASPLPPOEASNHQDCDDSMGYQYPTFLRVVQDGNSCAW 882
Qy 1170 CPWFRGCGKIDCGEDRAFTGNAYIAVDWDPALHLYRTSQERVVDEHSEVQERRAQ 1229
Db 883 CPWFRGCGKIDCGEDRAFTGNAYIAVDWDPALHLYRTSQERVVDEHSEVQERRAQ 942
Qy 1230 AEPINLDSCLRAFTSEBELGENEMYCSKCKTHCLATKCLDLWLPLPILIIHLKRFQFVN 1289
Db 943 AEPINLDSCLRAFTSEBELGENEMYCSKCKTHCLATKCLDLWLPLPILIIHLKRFQFVN 1002
Qy 1290 GRWIKSOKIVKFPRESFDPSPALVPRDPALCOHKPLTPQGDSELPRIAREVKKVDAQS 1349
Db 1003 GRWIKSOKIVKFPRESFDPSPALVPRDPALCOHKPLTPQGDSELPRIAREVKKVDAQS 1062
Qy 1350 SAGEDVLLSKSPSLSANIISPKGSPSSRSKGTSCPSKSNSSPNSSPRTILGRSKGRL 1409
Db 1063 SAGEDVLLSKSPSLSANIISPKGSPSSRSKGTSCPSKSNSSPNSSPRTILGRSKGRL 1122
Qy 1410 RLPOIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGSGQPELVTPODHEVALA 1469
Db 1123 RLPOIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGSGQPELVTPODHEVALA 1182
Qy 1470 NGELYEHACNGYNGOLGNHSEEDTDQREDTRIKPIYNLYASCHSGIILGGGHYTT 1529
Db 1183 NGELYEHACNGYNGOLGNHSEEDTDQREDTRIKPIYNLYASCHSGIILGGGHYTT 1242
Qy 1530 YAKPNCKWICYNDSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTDGKMAWDTSSM 1589
Db 1243 YAKPNCKWICYNDSCKELHPDEIDTDSAYILFYEQQIDYAOFLPKTDGKMAWDTSSM 1302
Qy 1590 DEDFESDYKCYVLQ 1604
Db 1303 DEDFESDYKCYVLQ 1317

RESULT 3

US-10-094-749-3045

; Sequence 3045, Application US/10094749

; Publication No. US20030219741A1

; GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOKUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 3045
LENGTH: 847
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-3045

Query Match 51.7%; Score 4428; DB 12; Length 847;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 831; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 19 RYTDVELKRLKDAFRTCGLSYTMGQHCFFIREVLGDGVPPKVAEVIYCSFGGTSGKLHFN 78
Db 16 RYTDVELKRLKDAFRTCGLSYTMGQHCFFIREVLGDGVPPKVAEVIYCSFGGTSGKLHFN 75
Qy 79 NLIVGLVLLTRGDKBEKAKYIFSLFSSSEGNVIREEMERMLHVVDGKVPDTRKCPSEG 138
Db 76 NLIVGLVLLTRGDKBEKAKYIFSLFSSSEGNVIREEMERMLHVVDGKVPDTRKCPSEG 135
Qy 139 EKVNYEKEFERNLFLNKDAFTFSRMLLSGVVYVTLTDDSDTPTFYQTLAGVTHLEESDIID 198
Db 136 EKVNYEKEFERNLFLNKDAFTFSRMLLSGVVYVTLTDDSDTPTFYQTLAGVTHLEESDIID 195
Qy 199 LEKRVWLLKAQSRTRGRPLETFGPIVSPPIRPSISEGLFNAPDENRDNHIDFKFISCGLS 258
Db 196 LEKRVWLLKAQSRTRGRPLETFGPIVSPPIRPSISEGLFNAPDENRDNHIDFKFISCGLS 255
Qy 259 ACCRFLAEROKCFKPFVDVDRDGVLSRVELDMVVALLEYVKONRTDDIPELHMDLSDI 318
Db 256 ACCRFLAEROKCFKPFVDVDRDGVLSRVELDMVVALLEYVKONRTDDIPELHMDLSDI 315
Qy 319 VEGILNAHDTTKMGHLLTLEDYQIWSVKVNLANEFLNLIFQVCHIVLGLRPAATPEEGQII 378
Db 316 VEGILNAHDTTKMGHLLTLEDYQIWSVKVNLANEFLNLIFQVCHIVLGLRPAATPEEGQII 375
Qy 379 RGWLERESRYGLQAGHNWFIISMQWQWKYVYKXNDANVPVIEPSSVLNGKYSFGTAAH 438
Db 376 RGWLERESRYGLQAGHNWFIISMQWQWKYVYKXNDANVPVIEPSSVLNGKYSFGTAAH 435
Qy 439 PMEQVEDRIGSSLSYVNTTEKFSNISTASEASTAGSFLYSATPGADVCFARQHNIS 498
Db 436 PMEQVEDRIGSSLSYVNTTEKFSNISTASEASTAGSFLYSATPGADVCFARQHNIS 495
Qy 499 DNNOCILGANGNILLHLNLPKPGAINQOPLVTPQEPVKATSLITLEGGRKRTPOLIHGRD 558
Db 496 DNNOCILGANGNILLHLNLPKPGAINQOPLVTPQEPVKATSLITLEGGRKRTPOLIHGRD 555

QY 559 YEMPEPVWALYHYGANLALPRVVKNSKTDIPELEFPRYLLFLRQOPATRTQGSNI 618
Db 556 YEMPEPVWALYHYGANLALPRVVKNSKTDIPELEFPRYLLFLRQOPATRTQGSNI 615
QY 619 WYNGVSPNAPLKRVLAYTGCFSRMOTIKIEIHEYLSORLRIKEEDMRWLWYNSNYLT 678
Db 616 WYNGVSPNAPLKRVLAYTGCFSRMOTIKIEIHEYLSORLRIKEEDMRWLWYNSNYLT 675
QY 679 LDDDEHKLLEYLKIODEQHLVIEVANKDMSPEMSFIANSKIDRHKVPTKATGLSN 738
Db 676 LDDDEHKLLEYLKIODEQHLVIEVANKDMSPEMSFIANSKIDRHEVPTKATGLSN 735
QY 739 LGNTCFMSSIOCVSNTQPTQYFISGRHLYELNETNPIGMKGHWAKCYGDLVQELWSGT 798
Db 736 LGNTCFMSSIOCVSNTQPTQYFISGRHLYELNETNPIGMKGHWAKCYGDLVQELWSGT 795
QY 799 QXNVAPLKLRTWITAKYAPRFNGFOQDSQELLAFLLDGLHEDLNRVHEKPYV 850
Db 796 QXNVAPLKLRTWITAKYAPRFNGFOQDSQELLAFLLDGLHEDLNRVHEKPYV 847

RESULT 4

US-10-274-639-14
; Sequence 14, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DELSEANE, Angelo M.; GANDHI, Ameena R.
; APPLICANT: RAFALIA, April J.A.; LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: BAUGH, Mariah R.; BOROWSKY, Mark L.
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
; APPLICANT: LAL, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 7476667CD1
US-10-274-639-14

Query Match 36.6%; Score 3138; DB 12; Length 582;
Best Local Similarity 100.0%; Pred. No. 9.5e-253;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1023 MFTLTITGDLPPPIIPNGMNTVVPQGTGKNTGMVNGHMPSPDPGPFYIIIVHVK 1082
Db 1 MFTLTITGDLPPPIIPNGMNTVVPQGTGKNTGMVNGHMPSPDPGPFYIIIVHVK 60
QY 1083 MMRTELYFSSQKNRPSLFGMPLIVPCTVHTTKKDYDAVMTQVSRLASPLPPOASNA 1142
Db 61 MMRTELYFSSQKNRPSLFGMPLIVPCTVHTTKKDYDAVMTQVSRLASPLPPOASNA 120
QY 1143 QDCDDSMGYQYFPTLVVQKDGNSCAWCPEYFCGCKIDCGEDRAFTGNVIAVDWDT 1202
Db 121 QDCDDSMGYQYFPTLVVQKDGNSCAWCPEYFCGCKIDCGEDRAFTGNVIAVDWDT 180
QY 1203 ALHLRYQTSQERVVDHESVESQSRRAQAPINLDSCLRAFTSEELGENEMYCKCKTH 1262
Db 181 ALHLRYQTSQERVVDHESVESQSRRAQAPINLDSCLRAFTSEELGENEMYCKCKTH 240
QY 1263 CLATKKLDLWRLPPLIHLK3FQFVNGEWIKSVKIVKPPRESFOPSFLVPRDPALCOH 1322
Db 241 CLATKKLDLWRLPPLIHLK3FQFVNGEWIKSVKIVKPPRESFOPSFLVPRDPALCOH 300
QY 1323 KPLTPQGBELSEPRILAREVVKVDAQSSAGEEDVLLSKSPSSLSANISSPKGSPSSSRK 1382
Db 301 KPLTPQGBELSEPRILAREVVKVDAQSSAGEEDVLLSKSPSSLSANISSPKGSPSSSRK 360
QY 1383 SGTSCESSKNSSPNSSPRLGRSKGRLRLPQTSKKNLSSKENLDASKENGAGQICELA 1442
Db 361 SGTSCESSKNSSPNSSPRLGRSKGRLRLPQTSKKNLSSKENLDASKENGAGQICELA 420
QY 1443 DALSRGHVJGSGOPELVTPQDHEVALANGFLYEACGNGYNGQLGHNSEEDSTDDQRE 1502
Db 421 DALSRGHVJGSGOPELVTPQDHEVALANGFLYEACGNGYNGQLGHNSEEDSTDDQRE 480
QY 1503 DTRIKPIYNLVIAISCHSGILGGHYVYVYAKNPNCKWYCYNDSSCKELHDEIDTDSAYIL 1562
Db 481 DTRIKPIYNLVIAISCHSGILGGHYVYVYAKNPNCKWYCYNDSSCKELHDEIDTDSAYIL 540
QY 1563 FYEQQIGDYAQLPKTDGKMDTSSMDEDFESDYKCYVLQ 1604
Db 541 FYEQQIGDYAQLPKTDGKMDTSSMDEDFESDYKCYVLQ 582

RESULT 5

US-09-764-875-762
; Sequence 762, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P2002
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 762
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-875-762

Query Match 31.7%; Score 2711; DB 12; Length 539;
Best Local Similarity 93.3%; Pred. No. 4e-217;
Matches 512; Conservative 9; Mismatches 6; Indels 22; Gaps 2;
QY 117 ERMHLVVDGKVPDTRKCFSEGEKYNVEKFRNWLFLNKDAFTSRWLLSGVYVITLDD 176
Db 7 ERMHLVVDGKVPDTRKCFSEGEKYNVEKFRNWLFLNKDAFTSRWLLSGVYVITLDD 66

```
QY 177 DPTFTYQTLAGVTHLEESDIIDLEKRYMLLKAQSRGTFDLETFGLVSPPIRPSLSEGL 236
Db 67 DPTFTYQTLAGVTHLEESDIIDLEKRYMLLKAQSRGTFDLETFGLVSPPIRPSLSEGL 126
QY 237 FNAFDENRNHIDFKIEISGLSACCRGPIAERQKCFKVFVDVDRDGVLSRVELRDMVVAL 296
Db 127 FNAFDENRNHIDFKIEISGLSACCRGPIAERQKCFKVFVDVDRDGVLSRVELRDMVVAL 186
QY 297 LEVWKDNRDIDIPELHMDLSDIVEGILNAHDTTKMGLHLETDYQIWSVKVLANEFNL 356
Db 187 LEVWKDNRDIDIPELHMDLSDIVEGILNAHDTTKMGLHLETDYQIWSVKVLANEFNL 246
QY 357 FOVCHIVLGLRPAATPEEGQIIRGMLERESRYGLOAGHWNFIISMOWQKWEYKYDAN 416
Db 247 FOVCHIVLGLRPAATPEEGQIIRGMLERESRYGLOAGHWNFIISMOWQKWEYKYDAN 306
QY 417 PVVIEPSSVLNGKYSFGTAAPHPMEQVEDRIGSSLSYVNTTEEKESDNISTASESETAG 476
Db 307 PVVIEPSSVLNGKYSFGTAAPHPMEQVEDRIGSSLSYVNTTEEKESDNISTASESETAG 366
QY 477 SGFLYSATPGADVCFARQHTSDNNQCILLGANGNILLHNPQKPGADINOPLVTOEPVK 536
Db 367 SGFLYSATPGADVCFARQHTSDNNQCILLGANGNILLHNPQKPGADINOPLVTOEPVK 426
QY 537 ATSLTEGGRLKRTPOLIHGRDYEMVPEVPWVRALYHWYGANLALPRPVIKSKTDIPELE 596
Db 427 ATSLTEGGRLKRTPOLIHGRDYEMVPEVPWVRALYHWYGANLALPRPVIKSKTDIPELE 486
QY 597 LFPFYLLFLRQOPATRTQOSNIWVNGNVPSPNAPLKRVLAYTGCFSRMOTIKETHEYLS 656
Db 487 LFPFYLLFLRQOPATRTQOSNIWVNGNVPSPNAPLKRVLAYTGCFSRMOTIKETHEYLS 525
QY 657 Q-RLRIKKEE 664
Db 526 RGNLRMQQE 534

RESULT 6
US-10-158-057-222
; Sequence 222, Application US/10158057
; Publication No. US20040014039A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ205C1
; CURRENT APPLICATION NUMBER: US/10/158,057
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 222
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (60)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (66)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (69)
```

```
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (76)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-158-057-222

Query Match 20.4%; Score 1750; DB 12; Length 401;
Best Local Similarity 88.1%; Pred. No. 4.6e-137;
Matches 334; Conservative 9; Mismatches 14; Indels 22; Gaps 2;

QY 287 VELRDMVVALLEVWKDNRDIDIPELHMDLSDIVEGILNAHDTTKMGLHLETDYQIWSVK 346
Db 39 VELRDMVVALLEVWKDNRDIDIPELHMDLSDIVEGILNAHDTTKMGLHLETDYQIWSVK 98
QY 347 VLANEFINLLFOVCHIVLGLRPAATPEEGQIIRGMLERESRYGLOAGHWNFIISMOWQ 406
Db 99 VLANEFINLLFOVCHIVLGLRPAATPEEGQIIRGMLERESRYGLOAGHWNFIISMOWQ 158
QY 407 WKYVYKDANPVVIEPSSVLNGKYSFGTAAPHPMEQVEDRIGSSLSYVNTTEEKESDNIS 466
Db 159 WKYVYKDANPVVIEPSSVLNGKYSFGTAAPHPMEQVEDRIGSSLSYVNTTEEKESDNIS 218
QY 467 TASEASETAGSGFLYSATPGADVCFARQHTSDNNQCILLGANGNILLHNPQKPGADIN 526
Db 219 TASEASETAGSGFLYSATPGADVCFARQHTSDNNQCILLGANGNILLHNPQKPGADIN 278
QY 527 QPLVTQEPVKATSLTEGGRLKRTPOLIHGRDYEMVPEVPWVRALYHWYGANLALPRPVIK 586
Db 279 QPLVTQEPVKATSLTEGGRLKRTPOLIHGRDYEMVPEVPWVRALYHWYGANLALPRPVIK 338
QY 587 NSKTDIPELELFPFYLLFLRQOPATRTQOSNIWVNGNVPSPNAPLKRVLAYTGCFSRMQ 646
Db 339 NSKTDIPELELFPFYLLFLRQOPATRTQOSNIWVNGNVPSPNAPLKRVLAYTGCFSRMQ 377
QY 647 TIXEIHYSQ-RLRIKKEE 664
Db 378 SLRWFQHLPRGNLRMQQE 396

RESULT 7
US-10-071-838-17
; Sequence 17, Application US/10071838
; Publication No. US20030044814A1
; GENERAL INFORMATION:
; APPLICANT: Li, Jing
; APPLICANT: Powers, Scott
; APPLICANT: Xiang, Phil
; APPLICANT: Peng, Yue
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: PRC17: An Amplified Cancer Gene
; FILE REFERENCE: 018781-007610US
; CURRENT APPLICATION NUMBER: US/10/071,838
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: US 60/267,615
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 786
; TYPE: PRT
```

ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human TRE-2/USP6
US-10-071-838-17

Query Match 16.4%; Score 1406; DB 15; Length 786;
Best Local Similarity 92.7%; Pred. No. 8.5e-108;
Matches 266; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 701 EVRNKMSWPEMSPFANSKIDRHKVPTEKGATGLNLTGNTCEMNSSIOCVNSTQTLTQ 760
DB 499 EVHNKMSWPEMSPFANSKIDRHKVPTEKGATGLNLTGNTCEMNSSIOCVNSTQTLTQ 558
QY 761 YFISGRHLYELNRTNPIGMKGMKACYGDLYQELWSGTQKNVAPLKLRTWITAKYAPFNG 820
DB 559 YFISGRHLYELNRTNPIGMKGMKACYGDLYQELWSGTQKNVAPLKLRTWITAKYAPFNG 618
QY 821 FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKSDGRPDWEVAABAWDNHLNRNRIIV 880
DB 619 FQOQDSQELLAFLLDGLHEDLNVRHEKPYVELKSDGRPDWEVAABAWDNHLNRNRIIV 678
QY 881 DLFGHQLRSQVKCKTCGHSVRFPPNFSLPLPMDSYMHLEITVTKLDGTPVRYGLRL 940
DB 679 DLFGHQLRSQVKCKTCGHSVRFPPNFSLPLPMDSYMHLEITVTKLDGTPVRYGLRL 738
QY 941 NMDEKYTGKQLSDLCGLNSEQILLAEVHGSNINFPQDNQKRLS 987
DB 739 NMDEKYTGKQLSDLCGLNSEQILLAEVHGSNINFPQDNQKRLS 785

RESULT 8
US-10-369-493-5738
Sequence 5738, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 5738
LENGTH: 1041
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-369-493-5738

Query Match 13.1%; Score 1124; DB 12; Length 1041;
Best Local Similarity 24.3%; Pred. No. 5.1e-84;
Matches 358; Conservative 188; Mismatches 436; Indels 490; Gaps 40;

QY 141 VNYEPRNWLFLNKDAFTSRWLLSGGVVYVTLTDDSDTFTYQTLAGVTHLEESDIIDLE 200
DB 16 ISAEADAKTYRFI-EYHFEFASGLSPKLTIPLNSYI-SFYEVMSYVTHLSVNEVIELE 73
QY 201 KRYWLKQASQGRFPLETFGLVPPPIRPSLSEGLFNAFENDRNHDHFKFISCGLSAC 260
DB 74 KVPATISDRA-VCKLNEEKWQALGCGFPDSYAERLFAVFDENRQDIDFRELVCITLSAL 132
QY 261 CRGPLAERQKFKFVDVDRDGVLSRVELRDMVALLEWKNRDTDDIPELMOLSDIVE 320
DB 133 CRGPLGRISQALARIWVDVDCDLLSDELSNM-----YKDL-----NVPEEHQVTV----- 178
QY 321 GILNAHDTTKMGHLLTLEDYQIWSVKM-VLANEFLLNLFQVCHIVLGLRPAATPEEGQIIR 379
DB 179 -----KSSNGKSALVDFGIWAQENKYNVYYSVALQIGHICLGLRPESRKMLQIWN 231

QY 380 GWLERESRYGLQAGHNWELISMQWQOMKEYKYKDANPVPVIEPSSVLNGKYSFCTAAHP 439
DB 232 EFEKASELPLS---EWNIVASGWAELRSFLADKNPNPDNSGI----- 274
QY 440 MBQVEDRIGSSLSYVNTTEKFSNDISTASEASETAGSGFLYSATPGADVCFARQHTSD 499
DB 275 -----KGTREDSWTS----- 284
QY 500 NNNQCLLGANGNILLHNPQKGAIDNQPLYTQEPVKATSLTLEGRLKRPQLLHGRDY 559
DB 285 -----KVACISAESARUK---PDLIPS-DY 305
QY 560 EMVPPVWREALYHYTCANLALPRPVIKN-----SKTDIPELELPYRLLFL---RQOPA 610
DB 306 IRVPVPLWEALRWHGCAITVDSQFTRKYLDGEFFDNKPALELYPLEILLGHDRKKSQ 365
QY 611 TTTQOSNIWNMGVNPSPNAPLKRVIATGTC---PSRQTIKEIHEIYLSQRLRIKEDARL 668
DB 366 DQTE-----NTP---RSLTSMWACAQVSRSMVTVDLALCKTELRLGDDARL 409
QY 669 WLYNSENY-LTLDDDEHKLLEYL-----KIQDEQHLVIEVRNKDMS-WPEEM----- 713
DB 410 WQVWKEEENGVLLDDGAQNLHQLYSSLGKTKVAKMKLLLEVRERGTVGWPPEELRASIS 469
QY 714 --SFIANSKIDRHKVPTEKGATGLNLTGNTCEMNSSIOCVNSTQTLTQYFISGRHLVEL 771
DB 470 GKQITAAASTLSSNAQLSGRPGAVGLVNYGNFCYRNAGIQCLARVSPLTQYFLDEDNLDAL 529
QY 772 NRTNPIGMKGMKAC---YGDLYQELWSGTQKNVAPLKLRTWITAKYAPFNGFQOQDSOE 828
DB 530 KRGNL---RGGDAETTIEYAKLLRENWAKKNIAF----- 563
QY 829 LLAFLLDGLHEDLNVRHEKPYVELKSD-GRPDWEVAABAWDNHLNRNRIIVDLFHGOL 887
DB 564 -----NDL-----IKKEADKEADEKAERSWTEYEKQNESLVTQLTGOL 605
QY 888 RSQVKCKTCGHSVRFPPNFSLPLPMDSYMHLEITVTKLDGTPVRYGLNLDKDYT 947
DB 606 RSKLICRTQSSSVPEPTSLSPIGFEDVDLYQYVVRDGRIPRRYIGFLSRDSKVG 665
QY 948 GLKKQLSDLCGLNSEQILLAEVHGSNINFPQDNQKRLSVSGFLCAFEIPVPVSPISAS 1007
DB 666 NLREVVAVSSGISMHLTIQ-----CMSSKGTL-----MSRSPNERS 702
QY 1008 SPTQTD--FSSSPSTNEMFTLTNGDLPRPIFIPNGMPTVVPCTGTEKNTFNGMNGHP 1065
DB 703 SNLRDELPLSSFPFGARLYAL----- 723
QY 1066 SLPDSPTG---YIIAVHRKMM-RTELYFLSSQKN-RPSLFQMPLIVPCTVHTRKDLVD 1120
DB 724 ELPEE--TGEDQWRVAMERKLQYNHEPYILGSTAGFIVSRFGLPLIVGLDEVTGKLYE 781
QY 1121 AVWIOVSRLASPLFPOEASNAHQDQDDSMGYQYPTTLRVVQKDGNSCAWCPYRRCRGCK 1180
DB 782 DVWYQHRFMEHSVNSSSRRAHDPCDE-NSGYPTTCLVDPNYEWCGCPALRFCRGP 840
QY 1181 IDCGEDRAFT-GNAYTAVDWDPTALHRYOTSOERVDEHESVRSRAQAAPINLDSCL 1239
DB 841 IRDESKVIFPANCPLAVDWLFIALVLRNHEQEQACEDDPVAETWSRHFAPSSLEHCL 900
QY 1240 RAFTSEELGENEMYCSKCTHCLATKKLDLWRPPLIIHLKRFQFV--NGRWIKSQK 1297
DB 901 EKFSCEPTEL--DAAIQDCRCKTMDKVTWIKLFPKYLIIHLKRFELREOGRMGKCKR 958
QY 1298 IVKFPRESDDPSAFVPRDPALCQHKPLTPQGDSELPILAREVKVKVDAQSSAGEDVL 1357
DB 959 TVNFPPLKGFDDPAFF----- 972
QY 1358 LSKSPSLSANITSSPKGSPSSSRKSGTSCPSKNSPNSPRTLGRSKGRLLRFPQIGSK 1417
DB 973 -----VDKPDGN----- 979

QY 1418 NKLSSKENLDASKENGAGQICELADALSRGHVLGSGQPELVTPQDHEVALANGFLYEHE 1477
Db 980 ----- 979
QY 1478 ACGNSYNGQLGNHSEEDSTDQREDTRIKPIYNYAIYASCHSGLGGHHYTYAKNPCK 1537
Db 980 -----TYECIALANHYQLSCGHFYAKSNEBK 1008
QY 1538 WYCYNDSSCKELHPDEIDTDSAYILFYEQGI 1569
Db 1009 WLLNDSCVRESEEVXQGNLYLFYERKDV 1040

RESULT 9

US-10-210-130-84

; Sequence 84, Application US/10210130

; Publication No. US20040014053A1

; GENERAL INFORMATION:

; APPLICANT: Zerhusen, Bryan D.

; APPLICANT: Patturajan, Meera

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Miller, Charles E.

; APPLICANT: Rieger, Daniel K.

; APPLICANT: Pena, Carol E.A.

; APPLICANT: Shinkets, Richard A.

; APPLICANT: Li, Li

; APPLICANT: Berghs, Constance

; APPLICANT: Zhong, Mei

; APPLICANT: Casman, Stacie J.

; APPLICANT: Voss, Edward Z.

; APPLICANT: Boldog, Ferenc L.

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Smithson, Glennda

; APPLICANT: Ji, Weizhen

; APPLICANT: Gorman, Linda

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Leite, Mario W.

; APPLICANT: Guo, Xiaojia Sasha

; APPLICANT: Anderson, David W.

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Gerlach, Valerie

; APPLICANT: Burgess, Catherine E.

; APPLICANT: Khramtsov, Nikolai V.

; APPLICANT: Ort, Tatiana

; APPLICANT: Ellerman, Karen

; APPLICANT: Rastelli, Luca

; APPLICANT: Agee, Michele L.

; APPLICANT: Chaudhuri, Amitabha

; APPLICANT: Chant, John S.

; APPLICANT: DiPippo, Vincent A.

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Eisen, Andrew J.

; APPLICANT: Gangolli, Bsha A.

; APPLICANT: Giot, Loic

; APPLICANT: Ooi, Chean Eng

; APPLICANT: Rothenberg, Mark E.

; APPLICANT: Spaderna, Steven K.

; APPLICANT: Hjal, Tord

; APPLICANT: Liu, Xiaohong

; APPLICANT: Taupier, Raymond J., Jr.

; APPLICANT: Catterton, Elna

; APPLICANT: Shenoy, Suresh G.

; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

; FILE REFERENCE: 21402-416C (Cura-716 SMT)

; CURRENT APPLICATION NUMBER: US/10/210,130

; CURRENT FILING DATE: 2002-08-01

; PRIOR APPLICATION NUMBER: 60/309,501

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: 60/316,508

; PRIOR FILING DATE: 2001-08-31

; PRIOR APPLICATION NUMBER: 60/354,655

; PRIOR FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: 60/310,291

; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/383,887
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/323,936
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: 60/381,039
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 84
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-210-130-84

Query Match 11.6%; Score 994; DB 12; Length 827;

Best Local Similarity 27.3%; Pred.No. 2.5e-73;

Matches 277; Conservative 136; Mismatches 287; Indels 314; Gaps 28;

QY 368 PATPEEGQIIRGLERESRYGLQAGHNWFIISMOWQWQWKEYKYDANVPVIEPSSVLN 427

Db 33 PGLDSQWRQIENGESGRER--PLRAGESWFLVKEKHWYKQWEAYVQ----- 75

QY 428 GGYKSEGTAAHPMEQVEDRIGSSLSYVNTTEKFSNISTASEASETAGSGFLYATPGA 487

Db 76 GG-----DQDSSTP----- 84

QY 488 DVCFAQHNTSDNNNOCLGANGNILLNLPQGAIDNOPLVTOBPVKATSLTLEGGR 547

Db 85 -----PCINNATLQDE-----INWRL 102

QY 548 KRTPLQIHGRDYEMVPEVVRALYHWYGANLALPRPVKNKSKTDIP---ELFLFRYLLF 604

Db 103 KE--GLVEGEDYVLLFAAAWHYLVSWYG--LBHGQPIERKVIETLQKVEVYVVELLL 158

QY 605 LRQQPATRTQQSNIVWVMGNVSPNAPLKRVLAYGCFSEMTQIKIEHYLSQRLRIK-E 663

Db 159 VRHN-----DLGK-----SHTVQSHSDISGLVLTARERFLVEPQ 194

QY 664 EDMRLWYNSENYLTLLDDDEHKLBYLKIQDBQHLVIEVRNKMDSWPSEMSFIANSKID 723

Db 195 EDTRLWAKNSEGLDLRYDTHITVLDAALETQQLIIMETRKKGDTWPSAQLHVMNNMSE 254

QY 724 RHK-VPTKGTGLNLGNTCEMNSSICVSNTPQLTOYFISGRHLYELNRTNPTGMKGH 782

Db 255 EDEDFKGQPGICGLTNLGNLTCFMSALQCLSNVPLQTEYFLNLCYELNFRNPLGMKGE 314

QY 783 MAKCYGDLVQELWSGTQKNVAPLKLRTIATKAPFNFGQQODSQELLAFLLDGLHEDLN 842

Db 315 IAEAYADLVKQAWSGHRSIVPHVFNKVGHFASQFLGYQQHDSQELLSFLLDGLHEDLN 374

QY 843 RVHEKPYVELKDSGRDPMEVAEAWDNHLNRNRISVVDLPHGQLRSOVKCTCGHISVR 902

Db 375 RYKCKEYVELCDAAGRPQOEVAEQNEKRENSDIVDTTFHGLFKSTLVCPDCGNVSVT 434

QY 903 FDPFNFLSLPLPMDSVHMLEITVVKLD-GTTVPVYGLRLNMDKVTGLKKQLSDCLGLNS 961

Db 435 PDPFCYLSVPLLSHKRVLEVFIPMDPRKPEQHRLVVPPKGGKISDLCLVALSKGTGSP 494

QY 962 EQLIAEAVHGSNIKPNPDQNKVLSVSGFLCAFEIPVPVPSFISASSPTQTDFFSSPSTN 1021

Db 495 ERMMVADVFSR-----FYKLYQLEERPLSSIL-----DRD 524

QY 1022 EMFTLTNG-----DLPRPIFIPNGMNTVVPCTEKNKFTNG-----MVNGH--NP 1065

Db 525 DIFVEVSGRIEAIEGSREDIVVPLRERTP-----ARDYNNSYGLMLFGHPLLV 576

Best Local Similarity 99.4%; Pred. No. 7.5e-66;
Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 117 ERMHVVVGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVYVTLTDDS 176
Db 1 ERMHVVVGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVYVTLTDDS 60
QY 177 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 236
Db 61 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 120
QY 237 FNAFDENRDNHIDFKFISCGLSACCRGFLAERQKFCFKVFDVDRDG 282
Db 121 FNAFDENRDNHIDFKFISCGLSACCRGFLAERQKFCFKVFDVDRDG 166

RESULT 13
US-09-764-875-1064
; Sequence 1064, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P202
; CURRENT APPLICATION NUMBER: US/09/764.875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1064
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (162)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-1064

Query Match 10.4%; Score 890; DB 12; Length 173;
Best Local Similarity 99.4%; Pred. No. 7.5e-66;
Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 117 ERMHVVVGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVYVTLTDDS 176
Db 1 ERMHVVVGKVPDILRKCFSEGEKYNVEKFNWFLFNKDAFTFRWLLSGGVYVTLTDDS 60
QY 177 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 236
Db 61 DTPFYQTLAGVTHLESDIIDLEKRYWLLKAQSGRTGREDLETFGPLVSPPIRPSLSEGL 120
QY 237 FNAFDENRDNHIDFKFISCGLSACCRGFLAERQKFCFKVFDVDRDG 282
Db 121 FNAFDENRDNHIDFKFISCGLSACCRGFLAERQKFCFKVFDVDRDG 166

RESULT 14
US-10-369-493-22797
; Sequence 22797, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369.493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22797
; LENGTH: 979
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22797

Query Match 9.2%; Score 787.5; DB 12; Length 979;
Best Local Similarity 22.9%; Pred. No. 6e-56;
Matches 283; Conservative 162; Mismatches 414; Indels 377; Gaps 37;
QY 432 SFGTAAPMEQVEDRIGSLSVNTTEKFSNISTASEASTAGSGLYSATPGADVCF 491
Db 19 SLSESSQSSSNVDDISQKSIISLGDAE--ISKVLSIAEQKQLIG----- 61
QY 492 ARQHTNSDNNOCILGANGNILLHNPQK-----PGAINQPLVTOEPVKATS 539
Db 62 ----ELVNNQPELELQGVNDYILSYSWYERLCSYLAEDGPPGPDVEDIA----- 108
QY 540 LLEGRLKRTPOLIHGRDYEMVPEPVRALYHWYG--ANLALPRVINKSKTDIPEL--E 596
Db 109 -DIETGTLK--PDLOEIDFTIISRDVMDLLVRWYGLKGPFPFRETVMLGSESHPLVVE 165
QY 597 LFPR-YLLFRQCPATRTQQSNINWYMGVSPNAPLKRVLAYTCFSRMQTIKEIHYL 655
Db 166 VVPIPFSLTLLSTNAVDANES-----HKPKKISL-----SKSTLEDLEGV 207
QY 656 SQRLAKSEDAEMLWYNSENVLTLLDDHDKLEYLKIQDBOHLVIEVRNKM----- 707
Db 208 KYTSLSPSQFELRWVDTDQPLHRTIDPS---SFIKINSKEIIDFLEKSKTLVELGMDSS 264
QY 708 -----SWPEE-----MSFI-----ANSKIDRHKVTEKGGATGLSNIGNTCF 744
Db 265 CSLVAECMNETWVDRLALRQFLIQORNNSSNEEQKEKVP---GTCGLSNIGNTCY 321
QY 745 MNSSTQCVSNTQPLTOYFISGRHLYELARTNPIGMKGMKACYDGLVOLMGSTQKNVAP 804
Db 322 MNSALQCLTHTRERLDRFTTSDEWKQVNESPLNGMGQGVAFASLIKSLYSPEHSEAP 381
QY 805 LKLRWTIAKAPRFGFQQDSQELLAFLLDGLHEDLNARVHEKPVV---ELKDSGRPDW 861
Db 382 RQKATIGKFNHISFLGYCQDSQEFLLAFLLDGLHEDLNRIYQKPYTSPDPDYEVEDEEK 441
QY 862 EVAABAWDNHLEARNKSIYVDFHGGQLRSQVCKTCGHISVRFPFNFLSLPLMDSMYHL 921
Db 442 NTAECBWRLLKRLNDSLIVDLFGQMYRSTLVCPCVNTVSIITFDPMDLTLPVVKQVMSH 501
QY 922 EITVIKLD-GTTPVRYGLRLNDE-----KYTGKQLSDLCGLNSEQILLAEVHGSN 973
Db 502 TVTFIPADTNLTPLAIEVVLSEKATIEDLVKYVAEKGSCSDY-----RKILVTETYKGR 556
QY 974 IKNF-PQDNQKVRLSVSG---FLCAFEIP-----VPVSPISASSPTQD--FSSSP 1018
Db 557 FYRFLTQLSKSLMEISEDEIYLYELRPEYDGDSDDLVPVYHISDDSTNSANSYMSR 616
QY 1019 STNEMFTL-TTNGDLPRPIFIPNGMNTVPCGTEKNTN-----GMVNGHMPSLPDSPP 1072
Db 617 DFGHPFVLQSDNEVTDASFISEKLKLYQQFTLLKLNKIDSLESLELGH-----EDQV 672
QY 1073 TGYIIAVHRKQMRTELYFLSSQKQR---PSLFGWELI-VPCTVHTRKKOLYDAWVQVS 1127
Db 673 QKGPLDVMDSHQSOTPLFEMRVFHDRFEKIPFGWMSVSNLPLTERDKO----- 722
QY 1128 RLASPLPPOEASNHAQCCDSDSGYQYPTFLRVQKDGNSCAWCPWYRFRCKGCKDCGE-D 1186
Db 723 -LESTVDPDLDAHSIEEDDSFQKVAPOGYPPEPSKSNEN-----TKLTAKEND 769
QY 1187 RAFIGNAVIADWDPTALHLYQT-----SQERVDEHESVQSRRAQAE 1232
Db 770 RLIIQGDLLVCWEPEKSYQFVFSVAPSPQWGRSLWLESKTLISDKDDSDSR---T 824
QY 1233 INLDSCLRAFTSEBELGENEMYCSKCKTHCLATKCLDMLRLPILLIHLKRFQFVNGRW 1292

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:36:36 ; Search time 28 seconds

(without alignments)
5509.092 Million cell updates/sec

Title: US-09-888-615-73

Perfect score: 8563

Sequence: 1 MGAKERIGFLSYEALRRV.....DTSSMDDEDFSDYKXVCVLQ 1604

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	4531	52.9	1089	2	S22158 transforming prote
2	1406	16.4	786	2	S22155 oncogene 1 (tre-2
3	1124	13.1	1041	2	T29010 hypothetical prote
4	1087	12.7	963	2	T09478 ubiquitin thiolest
5	1006	11.7	937	2	T04194 hypothetical prote
6	981.5	11.5	887	2	C86453 CDS protein F9L11.
7	980.5	11.5	794	2	I58376 hypothetical prote
8	957.5	11.3	928	2	T04192 hypothetical prote
9	920.5	10.7	914	2	T00757 probable ubiquitin
10	787.5	9.2	979	2	T41006 ubiquitin carboxyl
11	756.5	8.8	900	2	T33734 hypothetical prote
12	738.5	8.6	1254	2	S46636 hypothetical prote
13	693.5	8.1	849	2	T41085 probable ubiquitin
14	675	7.9	1095	2	T20528 hypothetical prote
15	371	4.3	926	2	S39344 deubiquinating enz
16	361.5	4.2	1083	2	T00750 ubiquitin-specific
17	330	3.9	805	2	S50277 ubiquitin-specific
18	330	3.9	961	2	H86181 hypothetical prote
19	313.5	3.7	268	2	T46902 hypothetical prote
20	303	3.5	438	2	T39772 hypothetical prote
21	259	3.0	445	2	I46225 ubiquitin carboxyl
22	249.5	2.9	1071	2	S48378 probable membrane
23	239	2.8	557	2	T50798 ubiquitin specific
24	211.5	2.5	717	2	S38177 SSV7 protein homol
25	203	2.4	457	2	T41069 ubiquitin carboxyl
26	203	2.4	555	2	T23531 hypothetical prote
27	200	2.3	526	2	JC6133 deubiquitinating e
28	190	2.2	754	2	S50601 hypothetical prote
29	184.5	2.2	875	2	T50182 ubiquitin-specific

30	183	2.1	785	2	T17309 probable ubiquitin
31	174.5	2.0	463	2	F85361 hypothetical prote
32	173.5	2.0	190	2	I21686 frequenin - Africa
33	167	2.0	190	2	A55686 neurocalcin - frui
34	167	2.0	190	2	S58303 related to neuroa
35	165	1.9	803	2	S45916 hypothetical prote
36	165	1.9	1292	2	A88225 protein T05H10.1
37	165	1.9	1292	2	T24559 hypothetical prote
38	164	1.9	199	2	S68838 guanylate cyclase-
39	162.5	1.9	1008	2	T05578 hypothetical prote
40	162	1.9	743	2	B84639 probable ubiquitin
41	161	1.9	193	2	JH0616 neurocalcin (clone
42	160.5	1.9	201	2	C55331 guanylate cyclase
43	160	1.9	688	2	S45803 hypothetical prote
44	159.5	1.9	371	2	T06097 ubiquitin-specific
45	159.5	1.9	471	2	S57591 hypothetical prote

ALIGNMENTS

RESULT 1

S22158:
transforming protein (clones 210 and 213) [imported] - human
N:Alternate names: oncogene 2 (tre-2 locus)
C:Species: Homo sapiens (man)
C:Date: 23-Apr-1993 #sequence revision 23-Apr-1993 #text_change 03-Nov-2000
R:Accession: S57868; S57875; S22156; S22158
R:Nakamura, T.; Hillova, J.; Mariage-Samson, R.; Onno, M.; Cannizzaro, L.;
Oncogene 7, 733-741, 1992
A:Title: A novel transcriptional unit of the tre oncogene widely expressed in human car
A:Reference number: S57867; MUID:92228503; PMID:1565468
A:Accession: S57868
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1089 <NAK1>
A:Cross-references: EMBL:X63547; NID:G37332; PIDN:CAA45111.1; PID:G37334
A:Experimental source: clone 213
A:Accession: S57875
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 504-1089 <NAK2>
A:Cross-references: EMBL:X63546; NID:G37329; PIDN:CAA45109.1; PID:G37331
A:Experimental source: clone 210
A>Note: submitted to the EMBL Data Library, December 1991

Query Match	52.9%	Score	4531;	DB	2;	Length	1089;
Best Local Similarity	93.7%	Pred.	No. 9.2e-284;				
Matches	851;	Conservative	22;	Mismatches	31;	Indels	4;
Gaps	1;						
Qy	701	EVNRKDMSPWPEMSPTANSSKIDRKVPTKGTGLSLNLTGTCFNMSSIQCVSNTQPLTQ	760				
Db	182	EVNRKDMSPWPEMSPTANSSKIDRKVPTKGTGLSLNLTGTCFNMSSIQCVSNTQPLTQ	241				
Qy	761	YFISGRHLYELNRTNPIGMKGMAKCYGDLVQELWSGQKXVAPLKLRTIAKYAPRNG	820				
Db	242	YFISGRHLYELNRTNPIGMKGMAKCYGDLVQELWSGQKXVAPLKLRTIAKYAPRNG	301				
Qy	821	FOQDSQELAFLLDLGHEDLNVRHEKPYVELKSDGSDGPDWEVAEAWDNLRRNSIVV	880				
Db	302	FOQDSQELAFLLDLGHEDLNVRHEKPYVELKSDGSDGPDWEVAEAWDNLRRNSIVV	361				
Qy	881	DLFHGQLASQVKCTCGHISVRFDPFNFSLPLPMDSYMHLEITVTKLDGTTVPVRYGLRL	940				
Db	362	DLFHGQLASQVKCTCGHISVRFDPFNFSLPLPMDSYMHLEITVTKLDGTTVPVRYGLRL	421				
Qy	941	NMDEKVTGLKQLSDLCGLNSEQILLAEVHGSNTKNFPQDNQKVLRSVSGFLCAFEIIVP	1000				
Db	422	NMDEKVTGLKQLSDLCGLNSEQILLAEVHGSNTKNFPQDNQKVLRSVSGFLCAFEIIVP	481				
Qy	1001	VSPISASPTQDFSSSPSTNEMFTLTNGDLPRPIFFIPNGMNTVVPDGGKKNFTNGMV	1060				
Db	482	VSPISASPTQDFSSSPSTNEMFTLTNGDLPRPIFFIPNGMNTVVPDGGKKNFTNGMV	541				

Db 410 WQVVKNEBGNVLDDGAQNLHOLYSSLGKTKKKNKMKLLEVRGRTGVWPEELRASUS 469
QY 714 --SFIANSKIDRHKVPTEKATGLNLTGNTCFMNSSIQCVSNTOPLTOYFISGRHLVEL 771
Db 470 GKQITAASTLSSNAQLSGRGVAGLVNNGFYRNAGIQCLARVSPLTQYFLDENLDAI 529
QY 772 NRTNPIGMKHWAKC---YGDVLOELWSGTQKVNAPLKLRTWTIAKYPFRNGFQQDQSOE 828
Db 530 KRGNL--RGDAAEETIEAKLEEMAAKKNIAF----- 563
QY 829 LLAFLLDGLHEDLNVRVHEKPYVELKDS--GRPDWEVAAEADNDHLNRNRSIVVDLFGHQL 887
Db 564 -----NDL-----IKKKEADEADEEKAERSWTEYEKQNESLTVQFTGQL 605
QY 888 RSQVKCTCGHISVRDPNPNFSLPLPMDSYMHELETVIKLDTTPVRYGLRLNMDKYT 947
Db 606 RSKLRCRTQSSSVPEPTSLSPICFEDVDLYQVIVVHRDGRIPRRYGFRLSRDSKVG 665
QY 948 GLKKQLSDLCGLNSEQIILAEVHGSNIKPNPDQNKVRLSVSGFLCAFEI PVPVSPISAS 1007
Db 666 NUREVAVSGISMSHLTQ-----CMSSKGTL-----MSRSPNHS 702
QY 1008 SPTQTD--FSSSPSTNEMFTLTNGDLPRPIFIPNGMPNTVVPCTGKNTFTNGVNGHMP 1065
Db 703 SNLRDELPLSSPFGARLYAL----- 723
QY 1066 SLPDSPTG--YIIAVERKMM-RTELYFLSSQKN-RPSLFGMPLIVPCTVTRKKDLYD 1120
Db 724 ELDES--TSDQWRVAHRKQJNHEPYILGSTAGFIVSRFGLPLIVGLDEEVTOKLYE 781
QY 1121 AVWQVSRLASPLPPOEASNAHQDQDSDSGYQYPTFLRVQDGNCSAFCWPCWYRCRCK 1180
Db 782 DVMYQHRFWERSVNSSSRAHDPCDE--NSGYPTLCLVDNENYWCQCPALRCRCP 840
QY 1181 IDCGEDRAFI--GNAYTAVDWDPTALHLRYOTQSRVVDHEHVSVEOSRAQAPINLDSCL 1239
Db 841 IRPDESKVIPANCPTAVDMLPIALYLYRNHSQEQACEDDPSVAETWGRHFAPSLEHCI 900
QY 1240 RAFTSEEELEGENMYCCKKCHLTKLKLWLRLPPLIILHLKRFQV--NGRWIKSOK 1297
Db 901 EKFSCPETL--DAAIQDCEKCKTWKDKWMTWKLPKYLIIHLKPEFLRQGRMGKCR 958
QY 1298 IVKFPRESDDPAFLVRDPALCOHKPLTPQGDSEPRILAREYKVKYDAOSSAGEEVL 1357
Db 959 TVNFFLKHDPAPP----- 972
QY 1358 LSKSPSSLANIIISPKGSPSSSRKSGTSCPSKSNSSPNSPRTLCRSKGRILRPLQISK 1417
Db 973 -----VDKPDGN----- 979
QY 1418 NKLSSKENLDASKENGAGQICELADALSRGHVGGSOPELVTPQDHEVALANGFLYEHE 1477
Db 980 ----- 979
QY 1478 ACGNGYNGQLGNHSEEDSTDQREDTRIKPIYNIYALISCHSGILGGHYTYAKNPCK 1537
Db 980 -----TYECIALANHYQLSCGHFIYAKSNEDK 1008
QY 1538 WYCYNDSSCKELHPDEIDTDSAYILFYEQGI 1569
Db 1009 WLLNDCSVREYSEEVKQGYLLFYERKDV 1040

RESULT 4
T09478
ubiquitin thiolesterase (EC 3.1.2.15) - human
N:Alternate names: ubiquitin carboxy-terminal esterase
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T09478
R:Gray, D.A.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z16684

A:Accession: T09478
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-963 <GRA>
A:Cross-references: EMBL:U20657; NID:g2459396; PID:g2459395
C:Genetics:
A:Gene: Unph
A:Map position: 3
C:Keywords: proto-oncogene; thiolester hydrolase

Query Match 12.78; Score 1087; DB 2; Length 963;
Best Local Similarity 23.94; Pred. No. 1.1e-61;
Matches 327; Conservative 168; Mismatches 313; Indels 560; Gaps 40;

QY 367 RPATPEEGQTIIRGLERESRYGLQAGHWNFIISMOWQOWKVEYKYDANPVVIEPSSVL 426
Db 10 RPDAAETQKSEL--GPLRRT--LQGAQWYLDLSRFKQWKYVGF----- 52
QY 427 NGGKYSFGTAHPMEQVEDRIGSSLSYVNTTEKFSNISTASEASETAGSGFLYSATPG 486
Db 53 SWDMYNYG----- 60
QY 487 ADVCFARQHNSTDNNNQCLLGANGNILLNLPQKGAIDNQLVTOEPVKATSLITLEGGR 546
Db 61 -----EHLNF-----PGIDNSGLFS--DP-----ESQT 82
QY 547 LKRTPOLIHRGDYEMVPEPVWRALYHYWGANTALPRPVIKNSKTDIPELELFP----- 600
Db 83 LKE--HLIDELDYLVPTAEANWLLNMYGC--VEGQQPIVRK-----VVEHGLFVKHCKVEV 135
QY 601 YLLFLRQCPATRTQQSNINWNGNVPSPNAPLAKRLAYTGC--FSRQTTKEIHEYLSQRL 659
Db 136 YLLEKL-----CENSDDPTNVL-----CHFSKADTIATIEKMRKLF 173
QY 660 RI--KEEDMRLWLYNSENLYLLDDEHKLLEYLKIODEHLVIEVRNKMDSWPE----- 712
Db 174 NIPARETRLWKNYMSNTVEQLSKLNTVQDAGLYQGVLIPEPQNEGTWPTQTLOSKS 233
QY 713 -----MSIAN-----SSKIDR-----HKVP 728
Db 234 STAPSNFTTSPKSSASPYSSVASLIANGDSTSTCGMHSSGVSRGGSGFSASYNQCEPP 293
QY 729 T--EKGATGLSLNLTGNTCFMNSSIQCVSNTOPLTOYFISGRHLYELNRTNPIGMKHWAK 785
Db 294 SSHIQFGLCGLNLTGNTCFMNSALQCLSNLTAPLDYFLKDEYEAENRDNPIGMKEIAE 353
QY 786 CYGDLVQELWSGTQKVNAPLKLRTWTIAKYPFRNGFQQDQSOELLAFLDLGLHEDLNVRH 845
Db 354 AYAEELIKOMWSGERDAHVPMSMFTQVGRFAPQPSGYQQDQSOELLAFLDLGLHEDLNVRK 413
QY 846 EKPYVELKDSGRPDWEVAAEADNDHLNRNRSIVVDLPHGQLRSQVKCKTCGHISVRDP 905
Db 414 KKPYLELKXANGRPDAVAKAEWENHRLRNDSVIVDTFHGLFKSLVCPKAVSVTFDP 473
QY 906 FNFSLPLPMDSYMHELETVIKLDTTPVRYGLRLNMDKYTGLKKQLSDLCGLNSEQI 964
Db 474 FCYLTPLPLKORVMEVFLPADPHCRPTQYRVTVPLMGAVSDLCCEALSRLSGIAAENM 533
QY 965 LLAFAVHGSNI--KNFPQDNQKRLSVSGFLCAFEI PVPVSPISASPTQTDFFSSSPSTNEM 1023
Db 534 VVADVYNHRFHKIFQMDGLNHPRDDIFVYEV-----CSTSVDSGSEC 577
QY 1024 FTLTNGDLPRPIFIPNGMPNTVVPCTGKNTFTNGVNGHMPSLPDSPTGTGTIAVHRKM 1083
Db 578 VTL-----PV----- 582
QY 1084 MRTELYFLSSQKNRPS-----LFGMPLIVPCTVH--TRKKDLYDAVWTVQVSR 1129
Db 583 -----YFERKSRPSSTSSALYQPLLSVPKHLTLESLYQAVCDRISRYVKQPLP 636
QY 1130 ---ASPLPPEASNAHQD---DDSMGYQYPTLRVVQKDGNSCAWCPWYRFGCGCKI 1181
Db 637 DEFGSSPLFPGACNRSNCSCEGEDEEHEHQEGKEQLSETEGS----- 680

QY 1182 DCGED-----RAF-----IGNAY-----I 1195
DB 681 --GEDEPNDSPTQKKIKQGPCPKRLFTPSLVNSYGTADINSLAADGKLLKLNRSSTL 738
QY 1196 AVDMDPTALHLRYQTSQBRVDEHESVQSRRAQAEPINLDSCLRAFTSEBELGENEMY 1255
DB 739 AMDMDRETRLLYYDSQSEAVEKHYVLMQPKKKTTVALRDCIELFTTMTLGEHDPWY 798
QY 1256 CSKCTHCLATKCLDLWLPPIILHLKRFQVNGRW-IKSQKIVKPPRESFDSAFLVP 1314
DB 799 CPNCKHQATKFKLWLPKILVVLHARFSY-ARYWRDKLDTVVEPPIRGLNMSF--- 854
QY 1315 RDPALCQHKPLTPQDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSLSANIISPK 1374
DB 855 ----- 854
QY 1375 GSPSSSRKSGTSCPPSSKNSSPNSSPRTLGRSKGRRLRPLQIGSKNKLSSKENLDAKENG 1434
DB 855 ----- 854
QY 1435 AGQICELADALSRGHVGLGSGQPELVTPODHEVALANGFLYEACGNGYNGQLGNHSEE 1494
DB 855 ---VCNLS----- 859
QY 1495 DSTDDQREDTRIPK-IYNLYAISCHSGILGGHYVYIAKNP-NCKWYCYNDSCKELHPD 1552
DB 860 -----ARPYVDLIAVNSHYGANGVGHYVAYAKNKLNGKYYFYDDNSVLASED 908
QY 1553 EIDTDSAYILFVEQOIGDYAQLPKTDKQVAD---TSSNDEDPESD 1596
DB 909 QIVTKAAVYLFYORDDDEFYK-TPSLSSSGSGDGGTRPSSQGGFGDD 955

RESULT 5
T04194
hypothetical protein T4F9.50 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04194
R:Bevan, M.; Wedler, H.; Wedler, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15260
A:Accession: T04194
A:Molecule type: DNA
A:Residues: 1-937 <BEV>
A:Cross-references: EMBL:AL049523
A:Experimental source: cultivar Columbia; BAC clone T4F9
C:Genetics:
A:Map position: 4
A:Introns: 51/2; 122/2; 170/3; 190/3; 250/3; 375/3; 422/3; 479/3; 565/3; 617/2; 795/2; 8
A:Note: T4F9.50

Query Match 11.7%; Score 1006; DB 2; Length 937;
Best Local Similarity 24.8%; Pred. No. 1.8e-56;
Matches 321; Conservative 153; Mismatches 336; Indels 482; Gaps 38;

QY 368 PATPEEGQILRGWLEERSRYLQAGHWNFIISOMWQOKEYKYDANVPVIFSSVLN 427
DB 18 PTPPEEKIV-SELITSEDNLKGSLNLYFVLSKGYTSWEKYVE----- 61
QY 428 GGYISFGTAHPMEQVEDRIGSSLSVNTTEKFSNDISTASEASETAGSGFLYSATPGA 487
DB 62 -----QSTKEYISGESSEAS----- 76
QY 488 DVCFAQRHTSDNNQCLLGANGNILLHNLPKPAINQAPLVTQEPVKASLTLEGRLL 547
DB 77 -----RPGFDNHDIISES-----DVNDPQL 98
QY 548 KRTPOLIGHRDYEMVPEPVYRALYHWYGANLALPRPVT-KNSKTDIPLELFPYLLFLR 606
DB 99 RRL--LMERVDIVLPQEWKRLVENVYSGGPPIERKLCOGFYTSYSEVYFICLMLTD 156

QY 607 QQPATFTQCSNIWANGVNPSPNAPLKVAVLTGCFSRMQTIKEIHEVLSORLRKEBDM 666
DB 157 GRDESRTV-----IRLQKQASIRELYEKVCAUTG-----VPQKFLMKDALRYEDF 204
QY 667 -----RLWLYNSENLYLLDDEPHK-LEVLIQIOEQLHLEVNRKOMSPE 711
DB 205 AALPHIDIFFYQAHIDYDFKRKNGLDLSYKSLERSSLHMDQDILLLEVDSGSSS-QS 263
QY 712 EMSFIANSKI-----DRHKV-----PT----- 729
DB 264 AMSSTNEALVLEPSSRSVVIAGGPTLSNGHSTSPNPSLPRTITSEDDGNSLSILCK 323
QY 730 -EKGA-TGLSNLGNCTCFMNSSIQCVSNTPQLPYFISGRHLYELNRTNPIGMKGMACY 787
DB 324 GEKGLAGLSNLGNTCFMNSALQCLAHTPIVEYFLQD--YSDDINRDNPLGMCGLAIATF 382
QY 788 GDVLQELWGSTQKNVAPLKLWTKIAYAPRFRNGQOQSQELLALFLDGLHEDLNRVHEK 847
DB 383 GDLKLLKLSGSRNSVAPRAFKTKLARFAPQFGYNGHDSQELLALFLDGLHEDLNKVRK 442
QY 848 PYVELKDSGRPDWEVAAEAWNDNHLRRNRSIVVDLFGQLRQSRQVCKTCGHSISVRDPFN 907
DB 443 PYIELKDSRDPDEVAEELWYHKAENDSVIVDCQGYKSTLVCPACGKISITFDPDM 502
QY 908 FLSPLPMDSYMHELTIVIKLGT-TPVRYGLRLNMDKEYTGLKQLSDLQGL-NSEQL 965
DB 503 YLSVPLPSTLTSMVTYVFCYDGLHPMTYVIVFKNGSIRDLTITAGTACLLAEDESLL 562
QY 966 LAEVHGSNTKNPQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTQTPSSSPSTNEMPT 1025
DB 563 LAEVDHKKIPKY-----FENPL-----DLSLSIKDDEHIV 592
QY 1026 LITNGDLPRPIPIPGMNTVVPQTEKNTNGMNGHMPSLPDSPTFTGYIITAVHRKMR 1085
DB 593 AVELNQMPP---GSGKAKLEILHGGQKRPILSVRGRDVKLFGTPFVTYV----- 639
QY 1086 TELYFLSSQKNRPSLFGMPLIIVPCTVHTKKDLYDAVWIQVSRLASPL---PPOEASNH 1141
DB 640 -----NTEPLSGADI-----DAV---LSRFLSPKHKVAPSKIENG 672
QY 1142 AQ-----DCDDSMGYQPTFLRVVQDGNSCAWCPWYRRCGCK 1180
DB 673 SENGHLPDATVDEASEILSSPDTEDDASRELSP-RIFLTDERGLNFKPLQ----- 723
QY 1181 IDCEDRAPIGNA-YIADWDPTALHLRYQTQSERVUDE-HESVEQSRAQAEPINLQSC 1238
DB 724 ---SESSI SLGTATRYLVEWN-EGEHERYDSSYLSOLPEVHKTSFSAKKTROESISLFS 779
QY 1239 LRAFTSEELGENEMYCYCKCKTHCLATKKLDLWLPPIILHLKRFQVNGRWIKSQ-- 1296
DB 780 LEAFLAEPLGPDMMWFCPSCKEHRQANKKLDLWKLPIILVFLHAKFTY--SRYLKKNKD 837
QY 1297 KIVKPPRESFDSAFLVPRDPALCQHKPLTPQDELSEPRILAREVKVDAQSSAGEEDV 1356
DB 838 TFFNVFVHDLDSLKYV----- 853
QY 1357 LLSKSPSLSANIISSPKGPSSSRKSGTSCPPSSKNSSPNSSPRTLGRSKGRLLPQIGS 1416
DB 854 ----- 853
QY 1417 KNKLSSENLDASKENGAGQICELADALSRGHVGLGSGQPELVTPODHEVALANGFLYEH 1476
DB 854 KNKNDQS-----YLYE- 864
QY 1477 EACGNGYNGQLGNHSEEDSTDQREDTRIKPIYNLYAISCHSGILGGHYVYIAK--NP 1534
DB 865 -----LYAVSNHYGGLGGHYIYAKLIDD 889
QY 1535 NCKWYCYNDSCKELHPDIDTDSAYILFVEQ 1566
DB 890 N-SWYHFDSDSHVSSVNESSEIKNSAAVLYFYRR 920

Db 10 ERETRLWNTKMSYTYEQSLKDLNTIQDAGLYQGVLVIEPQNEDEGWTPQSQSKSSTAP 69
QY 713 -----MSFTAN-----SSKIDR-----HKUPT--- 729
Db 70 SRNFTTSKPSASPYCSVASLIANGDSNTSSGMHSSGVSFGSFSASYNQOEAPSPHI 129
QY 730 EKGATGLSNLGNFCFMSNIGQVSNQPTQYPIGRHLYELNRPINPIGMKGMKAKCYGD 789
Db 130 QPGLCGLGNLGNFCFMSALQCLSNAPLTXEFLKDEYAEINRDNPLGKMGGEIAEYAE 189
QY 790 LVGELNSGCKVAPKLRWTIAKAPRNGFOQDSQELLAFDLGLHEDLNVRHVKPY 849
Db 190 LIQOMSGRDTHTVAPRFTQVGRFAPQSGVQOQDSQELLAFDLGLHEDLNVRHVKPY 249
QY 850 VELKSDGRPDWEAAEAMNHLRRNRISIVDLFHGQLRSQVKCKTCGHISVRFPFNFPL 909
Db 250 LEPKEPNRGPDAVAXEAMENHLRNDISIVDTFHGLFKSTLVCPCEKAVSTVFPFCYL 309
QY 910 SLPLPMSYMHLEITVILKLD-GTTPVRYGLRLNMDKTYTGLKKQLSDICGLNSQIILAE 968
Db 310 TLEPLKKORIMEVFLVPADPQCRPIQYRTVPLMGAISDLCEALSKLSGIAAENWVTD 369
QY 969 VHGSI-KNPPONQKVRLSVSGFLCAFELPVPVSPISASSPTQDTDFSSSPSTNEMFILT 1027
Db 370 VYNHREHKIFQMDGSLHTIPRDDIFVEV-----CNTSMGSGCITL- 412
QY 1028 TNGDLRPIPIPNMGNTVPCCTEKNTNGMNGMHPSLPSPPTGYIIAIVHKKMRT 1087
Db 413 -----PVVF-----REKK-----SRPSSASSGAV- 431
QY 1088 LYFLSSQKNRPSLFGMLIVPCIVH-TRKOLYDAYNIOVR-----LASPLP- 1135
Db 432 -----LYGQPLIVSPKPKLTLESIIQAVCDRISRVIKQPLPDEFLSPLEPG 479
QY 1136 -----QASNAHQD-----CDDSMG-----XQYPT 1156
Db 480 ACNGSRSSYEGDEEHEQEQLESEVSGGEDQDGDHSESAQVKVQPRHKLFT 539
QY 1157 LRUVQKDGNSCANCWPVRCGCKIDGEDRAFIGNAYIADVDPALHLRVTQSERVV 1216
Db 540 FSLV-----NSCGTADINSLATDGL-----LKNRSRTIAIDWDSSTRSLYDFDESEAC 590
QY 1217 DEHESVEQSRRAQEPINLDSCLRAFTSBEELGENEMYCSKCKTHCLATKKLDLWRLLP 1276
Db 591 EKHLNSQPQKKKAAVALRECIETFTMETLGEHDPWYCTCKKHQQAATKFDLWSLPK 550
QY 1277 ILIHLKRFQFVNGRW-IKQKI VTPRSPDPSPAFVLPDRPALCOHKPLTPQGBELSEP 1335
Db 651 ILVVHLKRFYSY-NRYWRDKLDTVVEFPFRALNMSEF----- 685
QY 1336 RILAREVKVYDAQSSAGEEDVLLSKGSPSLSANIISPKGSPSSSRKSGTSCPSSKNSSP 1395
Db 686 ----- 685
QY 1396 NSSPRTIGRSGKRLRPQIGSKNKLSSKENLDASKENGAGQICELADALSRHVLGSGQ 1455
Db 686 -----VC----- 687
QY 1456 PELVTPQDHEVALANGFLYEACNGSYNGQLGNHSEEDSTDQREDTRIKP-LYNLYA 1514
Db 688 -----DRSAPYVVDLIA 700
QY 1515 ISCHSGILGCGHYVYVAKNP-NCKWYCYNDSSCKELHPDEIDTDSAYILFYEQQGDIDYAO 1573
Db 701 VSNHYGAMGVGHYTAVALKRLNGKYYVFDSSVSLASEDQIVTKAAYVLFYQRRDECSS 760
QY 1574 -----FLPKTDGKKMADTSSMDEDFSDY 1597
Db 761 TSSLGSPFGDGGVKLSSHQGMGDEEAY 789

RESULT 8

T04192
hypothetical protein T4F9.30 - Arabidopsis thaliana
N:Alternate names: protein T4F9.30
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04192
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15260
A:Accession: T04192
A:Molecule type: DNA
A:Residues: 1-928 <BEV>
A:Cross-references: EMBL:AL049523
A:Experimental source: cultivar Columbia; BAC clone T4F9
C:Genetics:
A:Map position: 4
A:Interons: 51/2; 122/2; 161/3; 181/3; 214/3; 328/1; 401/3; 458/3; 544/3; 596/2; 774/2; 1
A:Note: T4F9.30

Query Match 11.3%; Score 967.5; DB 2; Length 928;
Best Local Similarity 24.3%; Pred. No. 5.5e-54;
Matches 320; Conservative 153; Mismatches 352; Indels 493; Gaps 38;
QY 359 VCHIVLGLRATPEEEGQIIRGHLERESRYGLOAGHNWFIISQWQWQWQKEYVYKIDANPV 418
Db 14 VCDL-----PFTPEERKIV-SELTSEEDNLKQGNLYFVISKRWYTSWQYVENSANE- 66
QY 419 VIEPSSVLNGGKYSFGTAHPMEQVEDRIGSSLSVYNTTEKFSNISTASEASETAGSG 478
Db 67 -----CSTGESSEA----- 75
QY 479 FLYSATPGADVCFARQHTSDNNNOCLLGANGNILLHNPQKPGALDNPQVTFQEPKAT 538
Db 76 -----PRPGIDNHDIIISDS----- 91
QY 539 SLTLEGGRKRLTPOLIHGRDYEMVPEVPMRVALYHWYGANLALPRPVTKSKTDIPELELF 598
Db 92 --DINDQLRL--LVGEGDYVLVPQVWKLVECG- -----PIE-- 128
QY 599 PRYLLFRQAPRATQOSNIWNGNVPSPNAPLKRVLAYTGCFSRMOTIKETHEYLQSQR 658
Db 129 -RKLIC--QGFYTRSYSEVY-----FLCLMLTDGRDESRTASIRELYEKVCAM 174
QY 659 LRKEEDMRWLNSENYLTLLDDEHK-LEYLKIQDBQHLVIEVRNKMSPWEEMSFA 717
Db 175 TGVQVEKAHIWDIFDKRKNGLLDPLSYKLEBSLESLHWDQDILVEVDGLSSSSQASSTG 234
QY 718 NSSKIDRHKVPTK----- 731
Db 235 NELAL-----VPLEPSRSIVTIAGPTLSNGHSTTSNPSLPFRITSEDDGRDSLILKGE 290
QY 732 --GATGLSNLGNFCFMSNIGQVSNQPTQYPIGRHLYE-----LNRTNP 776
Db 291 KGGIAGLSNLTGNTCFMNSALQCLAHPTPIVEYFLQDYMLDTPYFYLWVLTICILNILT 350
QY 777 IGMKGMKACVGDIVQVLSGQTKVAPLKLRTIAKAPRNGFOQDSQELLAFLLDG 836
Db 351 LNFQGEIAIAFGDLLKKLWSSGRNAVAPRAFKTLARFAPQPSGYNQCHDSQELLAFLLDG 410
QY 837 LHEDLNVRHVKPYVELKSDGRPDWEVAAEAMNHLRRNRISIVDLFHGQLRSQVKCKTC 896
Db 411 LHEDLNVRHVKPYVELKSDGRPDWEVAAEAMNHLRRNRISIVDLFHGQLRSQVKCKTC 470
QY 897 GHISVRPDPNFSLPLPMDSYMHLEITVILKLDGT-TPVRYGLRLNMDKTYTGLKQLSD 955
Db 471 GKISITDPPFWYLSVPLPSTLTRSMITTVFCGSRPLMPYTVIVPKQSIRDLITALGT 530
QY 956 LCGL-NSEQIILAEVHGSNKNFPQNKVRLSVSGFLCAFELPVPVSPISASSPTQDTDF 1014
Db 531 ACCLAEDESILLAEVYDHKIFRY-----FEIPLD-----SLSAIKDDEH 569
QY 1015 SSSPSTNEMFTLITNGDLPRPIPIPNMGNTVPCCTEKNTNGMNGMHPSLPSPPTGYIIAIVHKKMRT 1072

Db 802 VLVHLKRFYSRSMKLETFVNFPIHDLTKYV----- 837
QY 1337 ILAREVKVDAQSSAGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPSSKNSPN 1396
Db 838 ----- 837
QY 1397 SSPRTLGRKGRRLPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGGSP 1456
Db 838 -----ANKNL-----SQP 845
QY 1457 ELVTPQDHEVALANGFLYEACGNGYNGQLGNHSEEDSTDQREDTRIKPIYNYLAIS 1516
Db 846 QL-----YELYALT 854
QY 1517 CHSGILGGGHVYTYAK-NPNCKWYCYNDSSCKELHPDEITDTSVILFYEQQ 1567
Db 855 NHYGGMGSGHYTAHKLDDSRWYNFDDSHISHINEDDVKSGAAVLYFYRRK 906

RESULT 10
T41006
ubiquitin carboxyl-terminal hydrolase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T41006
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
Submitted to the EMBL Data Library, June 1998
A:Reference number: Z21963
A:Accession: T41006
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-979 <WOO>
A:Cross-references: EMBL:AL023776; PIDN:CAAL9303.1; GSPDB:GN00067; SPDB:SPCC1494.05c
A:Experimental source: strain 972h; cosmid ci494
C:Genetics:
A:Gene: SPDB:SPCC1494.05c
A:Map position: 2

Query Match 9.2%; Score 787.5; DB 2; Length 979;
Best Local Similarity 22.9%; Pred. No. 2.5e-42;
Matches 283; Conservative 162; Mismatches 414; Indels 377; Gaps 37;

QY 432 SFGTAHPMEQVEDRIGSSLSVYNTTEEFNSONISTASEASETAGSGFLYSATPGADVCF 491
Db 19 SLSEBSQSSNMDDISQKISLGDASE--ISKNLPSIAEQKQLIG----- 61
QY 492 ARQHTSDNNQCCLIGANGNILLHNPQK-----PGADNOLPLVTQBPVKATS 539
Db 62 ----ELVNNQPELELGQVDNYILSYSWYERLCSYLAEDGFFPGVDQEDIA----- 108
QY 540 LTLGGRLKRTPOLIRGRDYEMVPEVWRALYHWYG-ANLALPRVINKSKTDIDPEL--E 596
Db 109 -DLETGTLK--PDQBEIDFTIISRDVLLVWVGLKGPPEFRETVMLGSESHPLVVE 165
QY 597 LFPR-VLLFLRQOPARTQCSNIWNMGVSPFNAPLKRVLAYTCGFRMOTIKEIHEYL 655
Db 166 VYPIFSLTLLSTNAYDANES-----HKPKKISL-----SKSTLEDLLSGV 207
QY 656 SQRLRIKEDMRMLWLVNSNYLTLLDDEHKLLEYLKIODECHLVIEVRNKM----- 707
Db 208 KYTLSPSQFRLWRVDTPQLHRTIDPS---SFKINSKEIIDFLEKSKILVELGMDSS 264
QY 708 -----SWPEE-----MSFI-----ANSSKIDRHKVPTKEGATGLSNLGNTCF 744
Db 265 CSLVAECMINETWPDRLRLQFLIQORNQSSNEEQKEKVP---GTCSGLSNLGNTCY 321
QY 745 MSSICCVNTQPLTYQFYSGRHLVELARTNPIGKMGHAKCYGLDVELWGTQKQVAP 804
Db 322 MNSALQCLTHRELDRFFTSDEWKQVNESNPLMGGGQGVASIFASLIKLSYSPHESFAP 381
QY 805 LKLRWTIAYAPRFNGFOQDQSOELIAFLDGLDHEDLNRVHEKPYV---ELKDSGRPDW 861
Db 382 RQFKATIGFNHSGFLYGQDQSEFLAFLDGLDHEDLNRIYQKPYTSKPDLYEVDDEKIK 441

QY 862 EVAAEANDHLRNRSIVVDLPHGQLRSQVKCTCGHISVRPDPFNFELSLPLPMDSTMHL 921
Db 442 NTAECWRLKRLNDSLVLDLFOQMYRSTRLLVCPVNTVSTITDFPMDLTLLPLPVKQWWSH 501
QY 922 EITVIKLD-GTTPVRYGLRLNMDE-----KYTLKKQLSDCLGNSEQIILAEVHGSN 973
Db 502 TVTFIPADTNLTPLAIEVVLKESKAATIEDLVKVAESKGCSDY-----RKILVTEYKGR 556
QY 974 IKNF-PODNOKVRLSVSG-----FLCAFEIP-----VPVSPISASSPTQTD-FSSSP 1018
Db 557 FYRFLTLQLSKLLMEISEDEIYLYELERPEYDGSDDILVPVYHISDDSTNSANSYMSR 616
QY 1019 STNEMFTL-TTNGDLPRPIPIPNMGMENTVVPCTEKNFTN-----GMVNGHMPSLPDSPP 1072
Db 617 DFGHPFVLQSDNEVTDASFISEKILKYQQFTTLNKLKNIDLSLELGH-----EDEQV 672
QY 1073 TGYIIAHRKMRTELTYLSSQNR-----PSLFGMPLI-VPCTVTRKKDLYDAVWQVS 1127
Db 673 QKGFLDVDMDSQTPLEFMRVFDHREFEIKPTGWNNGSVSNLPLLTTERDKD----- 722
QY 1128 RLASPLPQEQASNAHACDDSMGVQVPTFLRVVQKDGNSCAWCPWYRFCRGCKIDCGE-D 1186
Db 723 -LESTVDPLDAHSIEDEDDSEPKDVAFGSYPEPSKSNEM-----TKLTAKEND 769
QY 1187 RAFIGNAYIAVDWDPTALHLRYQT-----SQERVVDEHESVQSRRAQAE 1232
Db 770 RLLTQGLLVCEWPEKSYQVFSVAPSSPQMGRLWLESKTLSDKKDSDSR-----T 824
QY 1233 INLDSCLRAFTSEELGENEMYCSKTHCLATKLDLWRLPPLIILHKKRFQVNGRW 1292
Db 825 ITLNDCLDEFEKTEQLGEEDFWYCTCKEFPQASKQMEIWRCEPILIFHLKRFSSERRFR 884
QY 1293 IKSQIKVFPRESFSAFLVPRDPALCOHKPLTPQGDSELPRIAREVKVYDAQSSAG 1352
Db 885 DKIDDLVEFP----- 894
QY 1353 EEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCPSSKNSPNSSPRTLGRSKGRRLP 1412
Db 895 ----- 894
QY 1413 QIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGGSPQLVTPQDHEVALANGF 1472
Db 895 -----INDLMSMTGSKYKLE----- 911
QY 1473 LYHEACGNGYNGQLGNHSEEDSTDQREDTRIKPIYNYLAISCHSGILGGGHVYTYAK 1532
Db 912 -----KENPKL--IYELIYAVDNHYGGLGGGHYTAFAK 941
QY 1533 NP-NCKWYCYNDSSCKELHPDEITDTSVILFYEQQ 1567
Db 942 NPDNGQFYCFDDSRVTPVCPPEETVTSAAVLLFYRRK 977

RESULT 11

T33734
hypothetical protein H34C03.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33734
R:Kalicki, J.; Gibson, A.
Submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid H34C03.
A:Reference number: Z21394
A:Accession: T33734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-900 <KAL>
A:Cross-references: EMBL:AF100662; PIDN:AAC68977.1; GSPDB:GN00022; CESP:H34C03.2
A:Experimental source: strain Bristol N2; clone H34C03
C:Genetics:
A:Gene: CESP:H34C03.2
A:Map position: 4

Db 244 -----SALNLRDLVEKSMNLFKADHLDVNAVDFKIFWSEGSDIATD 288
QY 674 ENYLTIILDD--EDHKLVELKI-----ODEQLVIEVR--NKDMSWP 710
Db 289 SNVSIFLNSSEYITPLQFLEIPKILLPDPENRDLKITSNPSDLVIEIKPIEGNHWP 348
QY 711 EWSFIANSKIDRHKVPTEKATGLSNLGNNTCFNNSIQCVSNTPQITQVFIQGRHLYE 770
Db 349 -----SNFYANKLEPASOTGLVNLGNNTCYNSALQCLVHIPQLRDYFLYDGYEDE 400
QY 771 LNRNTPIGMKHMAKCYGLVQELM-----SGQKXVA--PLKLRWTIAKYAPRNGFOOQ 824
Db 401 INEENPLGYGVAVAFSDLVQKLFQNRMSINORNAAPPPSMFKSTIGHFMSFGYMQQ 460
QY 825 DQOELAFLLDGLHEDLNVRVHEKPYVE---LKSDGRPDWEY---AAEAANDHLNRNR 877
Db 461 DSQEFLLFLDLSHEDLNRIIKKETEKPSPSPGDDVNDVNVKGLADTWEMELKRNCS 520
QY 878 IYVDLPHQQLRSQVCKTKCGHISVRPDPNPLSLPLPMDSYMHLEITVILKDGTTPIRYG 937
Db 521 VITDLPVGMKSTLYCPECONVSIITDPYNDVTLPLPVTVDWTKIKIPFN--SPPLLE 579
QY 938 LRLNMDKTYTGLKKQLSDLCGLNSQILLAEVHSGNIKNFPODNOKVRLSVSGFLCAPEI 997
Db 580 VELSKSSTVMDLKNVYGMKGLDPTLFCGEIFSQI---YVNVYESNESNAQFLTLQEL 635
QY 998 PVPSPISASSPTQDFSSPSSTNE-----MFTLTITNGDL 1032
Db 636 IRPADDVI-----FVELPVTNDNEVIVPVLNTRIBKGYKNAMLFVGPFFITLKEDEL 687
QY 1033 PRPIFIPNGMPTV-----PCTEKNTFNGMVNGHMPSL-----PDSPTGTG-- 1075
Db 688 NNPGALRMKLQNRVHLSGGYIPPEPVGNRTDFADA-----FELLVVKYIPDVEFEQYKD 742
QY 1076 -----IAVHRKMRTELTYFLSSQKRPSPFLGMPLI----- 1106
Db 743 ILQYTSIKVTRDKSGFFSIIKLSVEK-----EQFASNRTGPN--FWTPIQLNLDRATD 796
QY 1107 VPCTVHTRKKDLY-----DAWIOV-----SELASP-----LP 1134
Db 797 IDKLEDDVVDKIYNTSSLDVCAEGVLMQVDDREGTEGSEAKNFPKPGSGDEENKETVT 856
QY 1135 POASNAHQDCDDSMGY-----QVFFTLRVVQKDN 1165
Db 857 NNENVNTNDRDEDMELTDVVEEDASTEPELTDKPEALDKIKSITSTFFALLSN---N 912
QY 1166 SCACWNTYFRCGCKIDCGEDRAPIGNAYIADVDPALHRYQTSQBRVVDHESVEQS 1225
Db 913 DIIVCEW-----SELGSNEAFSDDK--IYNWENPA-----TLPNKL--ENAKLERS 955
QY 1226 RRAQAEPINLDSCLRAFTSEELGENEMYCYCKCKTHCLATKLDLWELPPLIILHLKRF 1285
Db 956 -NAKERTITLDDCLQLFSKPEILGLTDSWYCTCKEHRQATQQLWNTPDLIILHLKRF 1014
QY 1286 QFNGRWIKSQIKVFPRESFPFPAFLVRPDPALCOHKLPTQGDSELPILAREVKV 1345
Db 1015 ESQSFSDXIDATVAFPIITDLSRYVVKD----- 1045
QY 1346 DAQSSAGEEDVLLSKSPSSLSANIISSPKGSPSSSRKSGTSCPSKSNSSPRTLCRS 1405
Db 1046 -----DPRGL----- 1050
QY 1406 KGRLLRPLQIGSKNKLSSKENIDASKENGAGQICELADALSRGHVLGSGQPELVTPQDHE 1465
Db 1051 ----- 1050
QY 1466 VALANGFLYEACGNGYNGQLGNHSEBDSDDQEDTRIKPINLYAISCHSGILGGG 1525
Db 1051 -----YDLYAVDNHRYGGLGGG 1067
QY 1526 HYVTYAKN--PNCKWYCYNDSSCKELHPDEIDTDSAYILFY 1564
Db 1068 HVTAYVGNFADKNWYVFDSDRSRTETAPENSIAGSAYILFY 1107

RESULT 13

T41085

probable ubiquitin carboxyl-terminal hydrolase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T41085

R/McDougall, R.C.; Rajandream, M.A.; Batrell, B.G.; Zimmermann, W.; Wambutt, R.
submitted to the EMBL Data Library, August 1999

A/Reference number: Z21822

A/Accession: T41085

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-849 <MCD>

A/Cross-references: EMBL:AL109957; PIDN: CAB53084.1; GSPDB: GNC00068; SPDB: SPCC16A11.12C

A/Experimental source: strain 972h-; cosmid c16A11

C/Genetics:

A/Gene: SPDB: SPCC16A11.12C

A/Map position: 3

A/Introns: 107/2

Query Match

Best Local Similarity 8.1%; Score 693.5; DB 2; Length 849;

Matches 233; Conservative 160; Mismatches 321; Indels 379; Gaps 32;

QY 534 PVKATSLTLEGRLKRTPOLIHRDYMVPVPMVRYALYHYG-ANLALPRPV--IKNSKT 590

Db 70 PITQWELLDEKNEKELKHS--LEESIDYSIVSASLWHLVEMFGLGLEGLAERKVLVGLAAE 127

QY 591 DIPELELFP-RYLLFLRQOPATRTQOSNIWVMGNVP-SP-----NAPLKRVLAYTCPSR 644

Db 128 QKPFVDIYPIINFTHLVFDP-----INGENTSYSPLYQIDEPYHSDPEYAFSFR 177

QY 645 MOTIKIIBHYLSORLAIKE-EDMLRLYNSEN---YLTLLDDEHKL-----EY----- 689

Db 178 SDTLRLSLYKQVMEAFQISDGSFRLWLNKSNLSRFVSLSEFNQDPATALLSEYAVCMT 237

QY 690 ---LKTQDQHLVIEVRNKMSPWPEMSFIANSKIDRHKVPTEKGTGLSNLGNTCFNN 746

Db 238 IFIDIDADGS--LLLEPQHPENGWLSLSD-----SITQEQLNTINKEIGLCGLVNLGNSCYNN 291

QY 747 SSIQCYSNTQPTQVTFISGRHLYELNRTNPIGMKGMKACYGLDVQEL--WSGTQKNVAPL 805

Db 292 SALQCMHHTHEITKPLSDSYEKDINYNNPLGMGKVALSYASLLKMIHTADMHSVSPS 351

QY 806 KLRWTIAKYAPRNGPQQDQSOELAFLLDGLHEDLNVRVHEKPYVELKSDGRPD----- 860

Db 352 SPKFIICEFNTYPSGVRQDSQEFIAFFLDGLHEDLNRIQIKPYFE-----RPDLFDEH 405

QY 861 ----WEVAAEANDNHLNRNRSIVVDLPHQQLRSQVCKCTCGHISVRPDPFNLSLPLMD 916

Db 406 PLHVQRVANQCDIHRKNDSSIIVLFQGNYSKSTLECSICYOKSTAFDPFMTLTPPTS 465

QY 917 SYMHLEITVILKDG-TPVRYGLRLNMDKTYTGLKKQLSD---LCGLNSEQILLAEVHGS 972

Db 466 AKWRHKVVVVPFGTQSPVELYLELLMESTVILQMKFQATEKLQKMGLEGELTACDIYRG 525

QY 973 NIKNFPQDNQKVLRSVGLCAFEIPVPVSPISASPTQTDSSSPSTNEMFTLTNGDL 1032

Db 526 KVKVLKNDKTSKKI----- 541

QY 1033 PRPIFIPNGMPTVVPCTGTEKNTNGMVG--HMPSLPDSPPFTGYIIAHRKMRMTLYPL 1091

Db 542 -----HKWDHVLVYGSTANGLTIPVHGCKRPAMG----- 572

QY 1092 SSQKNRPSLFGHPLIVPCTVHTRKKD-----LYDAVWIVQVSLASPLPQ 1136

Db 573 SYGSN--DVFGFPL-----QLNYSRNLVNDLVKSEIVELYRYAGIDVAIGTLQGLKEM 626

QY 1137 EASNAHQDCDDSGVQYQPTLRVWOKDGNSCAWCPWYRCRCKIDCGEDRAFIGNAYIA 1196

Db 627 ESKAGWEIKIEVK---RFEIVBEE-----EIV 653

QY 1197 VMDPTALHLYQTSQ-ERVVDHESVEOSRRAQAPINLDSCLRAFTGEEELGENEMY 1255
Db 654 ID-DKTVIMCLWNDQYKELFYNCWEIPEKIQHMSITLLEDCLLEFSKPEQLDQDSWY 712
QY 1256 CSKCKTHCLATKLDLWRLPILIIHKKRQFVNG---RWIKSQKIVKPPRSPDPAFL 1312
Db 713 CPCKKAPRPAKLEIWRLEPKILVILHNRFSHGGLRERRRRKRDVWYVFDLNLKQEL 772
QY 1313 VPRDPALCQHKPLTPQDELSEPRILAREVKVYDAOSSAGEEDVLLSKSPSSISANI 1372
Db 773 SP----- 774
QY 1373 PKGSPSSSRKSGTSCPSKSNKSPNSPRTLGRSGKRLRLPQIGSKNKLSSKENLDASKE 1432
Db 775 ----- 774
QY 1433 NGAGQICELADALSRGHVLSGQPELVTPQDHEVALANGFLYEACNGYNGQLGNHS 1492
Db 775 -----FIKHEWLSQKS----- 787
QY 1493 EEDSTDQREDTRIKPIYNYLAISCHSGILGGHYVYAKNPCK-WYCYNDSSCKELHP 1551
Db 788 -----MLDYAVDNHGHGMSNGHYTAYARDASSQTFKFDTAICEIDP 832
QY 1552 DEITDTSAYILFY 1564
Db 833 EDIVTSSAYILFY 845

RESULT 14
T20528
hypothetical protein F07A11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20528
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20528
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1095 <WIL>
A:Cross-references: EMBL:Z66511; PIDN:CAA91317.1; GSPDB:GN000020; CESP:F07A11.4
A:Experimental source: Clone F07A11
C:Genetics:
A:Gene: CESP:F07A11.4
A:Map position: 2
A:Introns: 32/2; 202/2; 404/3; 587/3; 794/3; 964/1; 1004/3; 1080/2

Query Match 7.9%; Score 675; DB 2; Length 1095;
Best Local Similarity 28.8%; Pred. No. 5.4e-35;
Matches 197; Conservative 81; Mismatches 223; Indels 182; Gaps 18;

QY 723 DRHKV-PTKGGATGSLNGLTCFMNSSIQCVSNTQPLTQYFTISGRHLYELNRTNPIGMKG 781
Db 411 DREKVCVVEQGYTGLRNIGNTCFMNAVLMVNNIELREYFLRNHYQPEINETNPLSSEG 470
QY 782 HWAKCGDLVQELMSGTKQKVAPLKRWIAKYAPRNGFOQDSQELAFLLDGLHEDL 841
Db 471 RLAKAFADFMHQWSHQKAIETQIKNIVABKASQFANFAQDAHEFLSLFLDGLHEDV 530
QY 842 NRVEKPYVELKDSQCRPDWEAAAWNHLNRNRSIVVDLPHGQLRSQVCKCTGCHISV 901
Db 531 NRVKKPLGTGVESHGRHDLDSVNEAKNHLNRDSIFVDFHGQLKSHVQPCNDRVSI 590
QY 902 RDPPTNLSLPLPMSVMHLEITVTKDQTPVRYGLRNMDKEYTGLKKQLSDCLGLNS 961
Db 591 TDFDPVYLVPPFPKN-----KQSTDLI----- 612
QY 962 EQILLAEVHGSNIKFPDQDQ-----KYRLSVSGFLCAF-----EIPVPVSPISASP 1009
Db 613 -----FWPLENQTKPKLTVSYSTEGTVADFLSVSVSETVRVPTNLRACEA 658

QY 1010 TOTDP-----SSPSTNEMFTLTNGD-----LPRPIFIPNGMPT-- 1045
Db 659 ICHKFDKIYSLDMKVDIVSPDLRFVFTHDEMFEFNEIIVILHVLQRELY-----RKNTKH 714
QY 1046 -----VVPQGTBKNFTNGMVGNGHMPSLPDSPTGYIIAIVHRKMR 1085
Db 715 ACHECGNSKLKACBECYDAVYCSKEQVANWSTGGH-----R 753
QY 1086 TELYFLSQKRRPS-LFGMPLIVPCTVHTKKOILYDAWIVQVSLASPLPQEBASNAHQ- 1143
Db 754 EE-----CSKRPSETVGHPLIVSL---PRSLTYQHLRYVLEAKSRHTPPGESSSSSV 805
QY 1144 -----DCDDSMGYQYPTLRRVQKDGNSCAWCPCWYFCRCCKTDCGEDRAFINAY 1194
Db 806 SSTPRRSVAAPTKKQKMFIRKLSGQND-----FGHSISDDEACQLESQGGY 857
QY 1195 IAVDWDPTALLHLRYQTSQERVVDEHESVEOSRRAQAPIN-----LDSCIRA 1241
Db 858 VSVNWINQRNGKPYITTIENR-KOIVDVEKSRQWNRNMGYVYKSNNSNTPHILQMLDL 916
QY 1242 FTSEELGENEMYCYCKKTHCLATKLDLWRLPPLIIHLKRF-----QFVNGRWIK 1294
Db 917 FSETERLKPEESWTCTCKENVEATKKLYRUPPVLIIQLKRPVYTAFTYQSSMAERSK 976
QY 1295 SQKIVKFPRESFDPSPAFVPRDP 1317
Db 977 DTRSVYPLESLDMSPLAETSP 999

RESULT 15
S39344
deubiquitinating enzyme SSV7 - yeast (Saccharomyces cerevisiae)
N/Alternate names: deubiquitinating enzyme DOA4; protein D4270; protein YBRF926; protein
C/Species: Saccharomyces cerevisiae
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Jun-2000
C:Accession: S39344; S49824; S54053; S58849; S61758; S30807; S67885
R:Papa, F.R.; Hochstrasser, M.
Nature 366, 313-319, 1993
A:Title: The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of
A:Reference number: S39344; MUID:94067315; PMID:8247125
A:Accession: S39344
A:Molecule type: DNA
A:Residues: 1-926 <PAP>
A:Cross-references: GB:U02518; NID:G408456; PIDN:AAC48915.1; PID:G408457
R:Richards, C.; Harris, D.E.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49823
A:Accession: S49824
A:Molecule type: DNA
A:Residues: 1-926 <RIC>
A:Cross-references: EMBL:Z46796; NID:G577794; PIDN:CAA86791.1; PID:G577796
R:Hunt, S.; Bowman, S.; Harris, D.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54031
A:Accession: S54053
A:Molecule type: DNA
A:Residues: 1-926 <HUN>
A:Cross-references: EMBL:Z49209; NID:G798897; PIDN:CAA89098.1; PID:G798920
R:Brandt, P.; Otto, B.; Ramlow, S.; Blocker, H.
submitted to the EMBL Data Library, January 1995
A:Reference number: S58832
A:Accession: S58849
A:Molecule type: DNA
A:Residues: 36-926 <BRA>
A:Cross-references: EMBL:X84162; NID:G706817; PIDN:CAA58985.1; PID:G706835
R:Brandt, P.; Ramlow, S.; Otto, B.; Blocker, H.
Yeast 12, 85-90, 1996
A:Title: Nucleotide sequence analysis of a 32,500 bp region of the right arm of Sacchar
A:Reference number: S61741; MUID:96381250; PMID:8789263
A:Accession: S61758
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A;Residues: 36-926 <BRW>
A;Cross-references: EMBL:X84162; NID:g706817; PIDN:CAA58985.1; PID:g706835
R;Latterich, M.
submitted to the EMBL Data Library, January 1995
A;Reference number: S30807
A;Accession: S30807
A;Molecule type: DNA
A;Residues: 1-326, 'K', 328-344, 'F', 346-374, 'TASW', 379-382, 'I', 384-406, 'T', 408-542, 'FRS', 544-578, 'K', 580-600, 'K', 602-622, 'K', 624-644, 'K', 646-666, 'K', 668-688, 'K', 690-710, 'K', 712-732, 'K', 734-754, 'K', 756-776, 'K', 778-798, 'K', 800-820, 'K', 822-842, 'K', 844-864, 'K', 866-886, 'K', 888-908, 'K', 910-930, 'K', 932-952, 'K', 954-974, 'K', 976-996, 'K', 998-1018, 'K', 1020-1040, 'K', 1042-1062, 'K', 1064-1084, 'K', 1086-1106, 'K', 1108-1128, 'K', 1130-1150, 'K', 1152-1172, 'K', 1174-1194, 'K', 1196-1216, 'K', 1218-1238, 'K', 1240-1260, 'K', 1262-1282, 'K', 1284-1304, 'K', 1306-1326, 'K', 1328-1348, 'K', 1350-1370, 'K', 1372-1392, 'K', 1394-1414, 'K', 1416-1436, 'K', 1438-1458, 'K', 1460-1480, 'K', 1482-1502, 'K', 1504-1524, 'K', 1526-1546, 'K', 1548-1568, 'K', 1570-1590, 'K', 1592-1612, 'K', 1614-1634, 'K', 1636-1656, 'K', 1658-1678, 'K', 1680-1700, 'K', 1702-1722, 'K', 1724-1744, 'K', 1746-1766, 'K', 1768-1788, 'K', 1790-1810, 'K', 1812-1832, 'K', 1834-1854, 'K', 1856-1876, 'K', 1878-1898, 'K', 1900-1920, 'K', 1922-1942, 'K', 1944-1964, 'K', 1966-1986, 'K', 1988-2008, 'K', 2010-2030, 'K', 2032-2052, 'K', 2054-2074, 'K', 2076-2096, 'K', 2098-2118, 'K', 2120-2140, 'K', 2142-2162, 'K', 2164-2184, 'K', 2186-2206, 'K', 2208-2228, 'K', 2230-2250, 'K', 2252-2272, 'K', 2274-2294, 'K', 2296-2316, 'K', 2318-2338, 'K', 2340-2360, 'K', 2362-2382, 'K', 2384-2404, 'K', 2406-2426, 'K', 2428-2448, 'K', 2450-2470, 'K', 2472-2492, 'K', 2494-2514, 'K', 2516-2536, 'K', 2538-2558, 'K', 2560-2580, 'K', 2582-2602, 'K', 2604-2624, 'K', 2626-2646, 'K', 2648-2668, 'K', 2670-2690, 'K', 2692-2712, 'K', 2714-2734, 'K', 2736-2756, 'K', 2758-2778, 'K', 2780-2800, 'K', 2802-2822, 'K', 2824-2844, 'K', 2846-2866, 'K', 2868-2888, 'K', 2890-2910, 'K', 2912-2932, 'K', 2934-2954, 'K', 2956-2976, 'K', 2978-2998, 'K', 3000-3020, 'K', 3022-3042, 'K', 3044-3064, 'K', 3066-3086, 'K', 3088-3108, 'K', 3110-3130, 'K', 3132-3152, 'K', 3154-3174, 'K', 3176-3196, 'K', 3198-3218, 'K', 3220-3240, 'K', 3242-3262, 'K', 3264-3284, 'K', 3286-3306, 'K', 3308-3328, 'K', 3330-3350, 'K', 3352-3372, 'K', 3374-3394, 'K', 3396-3416, 'K', 3418-3438, 'K', 3440-3460, 'K', 3462-3482, 'K', 3484-3504, 'K', 3506-3526, 'K', 3528-3548, 'K', 3550-3570, 'K', 3572-3592, 'K', 3594-3614, 'K', 3616-3636, 'K', 3638-3658, 'K', 3660-3680, 'K', 3682-3702, 'K', 3704-3724, 'K', 3726-3746, 'K', 3748-3768, 'K', 3770-3790, 'K', 3792-3812, 'K', 3814-3834, 'K', 3836-3856, 'K', 3858-3878, 'K', 3880-3900, 'K', 3902-3922, 'K', 3924-3944, 'K', 3946-3966, 'K', 3968-3988, 'K', 3990-4010, 'K', 4012-4032, 'K', 4034-4054, 'K', 4056-4076, 'K', 4078-4098, 'K', 4100-4120, 'K', 4122-4142, 'K', 4144-4164, 'K', 4166-4186, 'K', 4188-4208, 'K', 4210-4230, 'K', 4232-4252, 'K', 4254-4274, 'K', 4276-4296, 'K', 4298-4318, 'K', 4320-4340, 'K', 4342-4362, 'K', 4364-4384, 'K', 4386-4406, 'K', 4408-4428, 'K', 4430-4450, 'K', 4452-4472, 'K', 4474-4494, 'K', 4496-4516, 'K', 4518-4538, 'K', 4540-4560, 'K', 4562-4582, 'K', 4584-4604, 'K', 4606-4626, 'K', 4628-4648, 'K', 4650-4670, 'K', 4672-4692, 'K', 4694-4714, 'K', 4716-4736, 'K', 4738-4758, 'K', 4760-4780, 'K', 4782-4802, 'K', 4804-4824, 'K', 4826-4846, 'K', 4848-4868, 'K', 4870-4890, 'K', 4892-4912, 'K', 4914-4934, 'K', 4936-4956, 'K', 4958-4978, 'K', 4980-5000, 'K', 5002-5022, 'K', 5024-5044, 'K', 5046-5066, 'K', 5068-5088, 'K', 5090-5110, 'K', 5112-5132, 'K', 5134-5154, 'K', 5156-5176, 'K', 5178-5198, 'K', 5200-5220, 'K', 5222-5242, 'K', 5244-5264, 'K', 5266-5286, 'K', 5288-5308, 'K', 5310-5330, 'K', 5332-5352, 'K', 5354-5374, 'K', 5376-5396, 'K', 5398-5418, 'K', 5420-5440, 'K', 5442-5462, 'K', 5464-5484, 'K', 5486-5506, 'K', 5508-5528, 'K', 5530-5550, 'K', 5552-5572, 'K', 5574-5594, 'K', 5596-5616, 'K', 5618-5638, 'K', 5640-5660, 'K', 5662-5682, 'K', 5684-5704, 'K', 5706-5726, 'K', 5728-5748, 'K', 5750-5770, 'K', 5772-5792, 'K', 5794-5814, 'K', 5816-5836, 'K', 5838-5858, 'K', 5860-5880, 'K', 5882-5902, 'K', 5904-5924, 'K', 5926-5946, 'K', 5948-5968, 'K', 5970-5990, 'K', 5992-6012, 'K', 6014-6034, 'K', 6036-6056, 'K', 6058-6078, 'K', 6080-6100, 'K', 6102-6122, 'K', 6124-6144, 'K', 6146-6166, 'K', 6168-6188, 'K', 6190-6210, 'K', 6212-6232, 'K', 6234-6254, 'K', 6256-6276, 'K', 6278-6298, 'K', 6300-6320, 'K', 6322-6342, 'K', 6344-6364, 'K', 6366-6386, 'K', 6388-6408, 'K', 6410-6430, 'K', 6432-6452, 'K', 6454-6474, 'K', 6476-6496, 'K', 6498-6518, 'K', 6520-6540, 'K', 6542-6562, 'K', 6564-6584, 'K', 6586-6606, 'K', 6608-6628, 'K', 6630-6650, 'K', 6652-6672, 'K', 6674-6694, 'K', 6696-6716, 'K', 6718-6738, 'K', 6740-6760, 'K', 6762-6782, 'K', 6784-6804, 'K', 6806-6826, 'K', 6828-6848, 'K', 6850-6870, 'K', 6872-6892, 'K', 6894-6914, 'K', 6916-6936, 'K', 6938-6958, 'K', 6960-6980, 'K', 6982-7002, 'K', 7004-7024, 'K', 7026-7046, 'K', 7048-7068, 'K', 7070-7090, 'K', 7092-7112, 'K', 7114-7134, 'K', 7136-7156, 'K', 7158-7178, 'K', 7180-7200, 'K', 7202-7222, 'K', 7224-7244, 'K', 7246-7266, 'K', 7268-7288, 'K', 7290-7310, 'K', 7312-7332, 'K', 7334-7354, 'K', 7356-7376, 'K', 7378-7398, 'K', 7400-7420, 'K', 7422-7442, 'K', 7444-7464, 'K', 7466-7486, 'K', 7488-7508, 'K', 7510-7530, 'K', 7532-7552, 'K', 7554-7574, 'K', 7576-7596, 'K', 7598-7618, 'K', 7620-7640, 'K', 7642-7662, 'K', 7664-7684, 'K', 7686-7706, 'K', 7708-7728, 'K', 7730-7750, 'K', 7752-7772, 'K', 7774-7794, 'K', 7796-7816, 'K', 7818-7838, 'K', 7840-7860, 'K', 7862-7882, 'K', 7884-7904, 'K', 7906-7926, 'K', 7928-7948, 'K', 7950-7970, 'K', 7972-7992, 'K', 7994-8014, 'K', 8016-8036, 'K', 8038-8058, 'K', 8060-8080, 'K', 8082-8102, 'K', 8104-8124, 'K', 8126-8146, 'K', 8148-8168, 'K', 8170-8190, 'K', 8192-8212, 'K', 8214-8234, 'K', 8236-8256, 'K', 8258-8278, 'K', 8280-8300, 'K', 8302-8322, 'K', 8324-8344, 'K', 8346-8366, 'K', 8368-8388, 'K', 8390-8410, 'K', 8412-8432, 'K', 8434-8454, 'K', 8456-8476, 'K', 8478-8498, 'K', 8500-8520, 'K', 8522-8542, 'K', 8544-8564, 'K', 8566-8586, 'K', 8588-8608, 'K', 8610-8630, 'K', 8632-8652, 'K', 8654-8674, 'K', 8676-8696, 'K', 8698-8718, 'K', 8720-8740, 'K', 8742-8762, 'K', 8764-8784, 'K', 8786-8806, 'K', 8808-8828, 'K', 8830-8850, 'K', 8852-8872, 'K', 8874-8894, 'K', 8896-8916, 'K', 8918-8938, 'K', 8940-8960, 'K', 8962-8982, 'K', 8984-9004, 'K', 9006-9026, 'K', 9028-9048, 'K', 9050-9070, 'K', 9072-9092, 'K', 9094-9114, 'K', 9116-9136, 'K', 9138-9158, 'K', 9160-9180, 'K', 9182-9202, 'K', 9204-9224, 'K', 9226-9246, 'K', 9248-9268, 'K', 9270-9290, 'K', 9292-9312, 'K', 9314-9334, 'K', 9336-9356, 'K', 9358-9378, 'K', 9380-9400, 'K', 9402-9422, 'K', 9424-9444, 'K', 9446-9466, 'K', 9468-9488, 'K', 9490-9510, 'K', 9512-9532, 'K', 9534-9554, 'K', 9556-9576, 'K', 9578-9598, 'K', 9600-9620, 'K', 9622-9642, 'K', 9644-9664, 'K', 9666-9686, 'K', 9688-9708, 'K', 9710-9730, 'K', 9732-9752, 'K', 9754-9774, 'K', 9776-9796, 'K', 9798-9818, 'K', 9820-9840, 'K', 9842-9862, 'K', 9864-9884, 'K', 9886-9906, 'K', 9908-9928, 'K', 9930-9950, 'K', 9952-9972, 'K', 9974-9994, 'K', 9996-10000, 'K', 10002-10022, 'K', 10024-10044, 'K', 10046-10066, 'K', 10068-10088, 'K', 10090-10110, 'K', 10112-10132, 'K', 10134-10154, 'K', 10156-10176, 'K', 10178-10198, 'K', 10200-10220, 'K', 10222-10242, 'K', 10244-10264, 'K', 10266-10286, 'K', 10288-10308, 'K', 10310-10330, 'K', 10332-10352, 'K', 10354-10374, 'K', 10376-10396, 'K', 10398-10418, 'K', 10420-10440, 'K', 10442-10462, 'K', 10464-10484, 'K', 10486-10506, 'K', 10508-10528, 'K', 10530-10550, 'K', 10552-10572, 'K', 10574-10594, 'K', 10596-10616, 'K', 10618-10638, 'K', 10640-10660, 'K', 10662-10682, 'K', 10684-10704, 'K', 10706-10726, 'K', 10728-10748, 'K', 10750-10770, 'K', 10772-10792, 'K', 10794-10814, 'K', 10816-10836, 'K', 10838-10858, 'K', 10860-10880, 'K', 10882-10902, 'K', 10904-10924, 'K', 10926-10946, 'K', 10948-10968, 'K', 10970-10990, 'K', 10992-11012, 'K', 11014-11034, 'K', 11036-11056, 'K', 11058-11078, 'K', 11080-11100, 'K', 11102-11122, 'K', 11124-11144, 'K', 11146-11166, 'K', 11168-11188, 'K', 11190-11210, 'K', 11212-11232, 'K', 11234-11254, 'K', 11256-11276, 'K', 11278-11298, 'K', 11300-11320, 'K', 11322-11342, 'K', 11344-11364, 'K', 11366-11386, 'K', 11388-11408, 'K', 11410-11430, 'K', 11432-11452, 'K', 11454-11474, 'K', 11476-11496, 'K', 11498-11518, 'K', 11520-11540, 'K', 11542-11562, 'K', 11564-11584, 'K', 11586-11606, 'K', 11608-11628, 'K', 11630-11650, 'K', 11652-11672, 'K', 11674-11694, 'K', 11696-11716, 'K', 11718-11738, 'K', 11740-11760, 'K', 11762-11782, 'K', 11784-11804, 'K', 11806-11826, 'K', 11828-11848, 'K', 11850-11870, 'K', 11872-11892, 'K', 11894-11914, 'K', 11916-11936, 'K', 11938-11958, 'K', 11960-11980, 'K', 11982-12002, 'K', 12004-12024, 'K', 12026-12046, 'K', 12048-12068, 'K', 12070-12090, 'K', 12092-12112, 'K', 12114-12134, 'K', 12136-12156, 'K', 12158-12178, 'K', 12180-12200, 'K', 12202-12222, 'K', 12224-12244, 'K', 12246-12266, 'K', 12268-12288, 'K', 12290-12310, 'K', 12312-12332, 'K', 12334-12354, 'K', 12356-12376, 'K', 12378-12398, 'K', 12400-12420, 'K', 12422-12442, 'K', 12444-12464, 'K', 12466-12486, 'K', 12488-12508, 'K', 12510-12530, 'K', 12532-12552, 'K', 12554-12574, 'K', 12576-12596, 'K', 12598-12618, 'K', 12620-12640, 'K', 12642-12662, 'K', 12664-12684, 'K', 12686-12706, 'K', 12708-12728, 'K', 12730-12750, 'K', 12752-12772, 'K', 12774-12794, 'K', 12796-12816, 'K', 12818-12838, 'K', 12840-12860, 'K', 12862-12882, 'K', 12884-12904, 'K', 12906-12926, 'K', 12928-12948, 'K', 12950-12970, 'K', 12972-12992, 'K', 12994-13014, 'K', 13016-13036, 'K', 13038-13058, 'K', 13060-13080, 'K', 13082-13102, 'K', 13104-13124, 'K', 13126-13146, 'K', 13148-13168, 'K', 13170-13190, 'K', 13192-13212, 'K', 13214-13234, 'K', 13236-13256, 'K', 13258-13278, 'K', 13280-13300, 'K', 13302-13322, 'K', 13324-13344, 'K', 13346-13366, 'K', 13368-13388, 'K', 13390-13410, 'K', 13412-13432, 'K', 13434-13454, 'K', 13456-13476, 'K', 13478-13498, 'K', 13500-13520, 'K', 13522-13542, 'K', 13544-13564, 'K', 13566-13586, 'K', 13588-13608, 'K', 13610-13630, 'K', 13632-13652, 'K', 13654-13674, 'K', 13676-13696, 'K', 13698-13718, 'K', 13720-13740, 'K', 13742-13762, 'K', 13764-13784, 'K', 13786-13806, 'K', 13808-13828, 'K', 13830-13850, 'K', 13852-13872, 'K', 13874-13894, 'K', 13896-13916, 'K', 13918-13938, 'K', 13940-13960, 'K', 13962-13982, 'K', 13984-14004, 'K', 14006-14026, 'K', 14028-14048, 'K', 14050-14070, 'K', 14072-14092, 'K', 14094-14114, 'K', 14116-14136, 'K', 14138-14158, 'K', 14160-14180, 'K', 14182-14202, 'K', 14204-14224, 'K', 14226-14246, 'K', 14248-14268, 'K', 14270-14290, 'K', 14292-14312, 'K', 14314-14334, 'K', 14336-14356, 'K', 14358-14378, 'K', 14380-14400, 'K', 14402-14422, 'K', 14424-14444, 'K', 14446-14466, 'K', 14468-14488, 'K', 14490-14510, 'K', 14512-14532, 'K', 14534-14554, 'K', 14556-14576, 'K', 14578-14598, 'K', 14600-14620, 'K', 14622-14642, 'K', 14644-14664, 'K', 14666-14686, 'K', 14688-14708, 'K', 14710-14730, 'K', 14732-14752, 'K', 14754-14774, 'K', 14776-14796, 'K', 14798-14818, 'K', 14820-14840, 'K', 14842-14862, 'K', 14864-14884, 'K', 14886-14906, 'K', 14908-14928, 'K', 14930-14950, 'K', 14952-14972, 'K', 14974-14994, 'K', 14996-15016, 'K', 15018-15038, 'K', 15040-15060, 'K', 15062-15082, 'K', 15084-15104, 'K', 15106-15126, 'K', 15128-15148, 'K', 15150-15170, 'K', 15172-15192, 'K', 15194-15214, 'K', 15216-15236, 'K', 15238-15258, 'K', 15260-15280, 'K', 15282-15302, 'K', 15304-15324, 'K', 15326-15346, 'K', 15348-15368, 'K', 15370-15390, 'K', 15392-15412, 'K', 15414-15434, 'K', 15436-15456, 'K', 15458-15478, 'K', 15480-15500, 'K', 15502-15522, 'K', 15524-15544, 'K', 15546-15566, 'K', 15568-15588, 'K', 15590-15610, 'K', 15612-15632, 'K', 15634-15654, 'K', 15656-15676, 'K', 15678-15698, 'K', 15700-15720, 'K', 15722-15742, 'K', 15744-15764, 'K', 15766-15786, 'K', 15788-15808, 'K', 15810-15830, 'K', 15832-15852, 'K', 15854-15874, 'K', 15876-15896, 'K', 15898-15918, 'K', 15920-15940, 'K', 15942-15962, 'K', 15964-15984, 'K', 15986-16006, 'K', 16008-16028, 'K', 16030-16050, 'K', 16052-16072, 'K', 16074-16094, 'K', 16096-16116, 'K', 16118-16138, 'K', 16140-16160, 'K', 16162-16182, 'K', 16184-16204, 'K', 16206-16226, 'K', 16228-16248, 'K', 16250-16270, 'K', 16272-16292, 'K', 16294-16314, 'K', 16316-16336, 'K', 16338-16358, 'K', 16360-16380, 'K', 16382-16402, 'K', 16404-16424, 'K', 16426-16446, 'K', 16448-16468, 'K', 16470-16490, 'K', 16492-16512, 'K', 16514-16534, 'K', 16536-16556, 'K', 16558-16578, 'K', 16580-16600, 'K', 16602-16622, 'K', 16624-16644, 'K', 16646-16666, 'K', 16668-16688, 'K', 16690-16710, 'K', 16712-16732, 'K', 16734-16754, 'K', 16756-16776, 'K', 16778-16798, 'K', 16800-16820, 'K', 16822-16842, 'K', 16844-16864, 'K', 16866-16886, 'K', 16888-16908, 'K', 16910-16930, 'K', 16932-16952, 'K', 16954-16974, 'K', 16976-16996, 'K', 16998-17018, 'K', 17020-17040, 'K', 17042-17062, 'K', 17064-17084, 'K', 17086-17106, 'K', 17108-17128, 'K', 17130-17150, 'K', 17152-17172, 'K', 17174-17194, 'K', 17196-17216, 'K', 17218-17238, 'K', 17240-17260, 'K', 17262-17282, 'K', 17284-17304, 'K', 17306-17326, 'K', 17328-17348, 'K', 17350-17370, 'K', 17372-17392, 'K', 17394-17414, 'K', 17416-17436, 'K', 17438-17458, 'K', 17460-17480, 'K', 17482-17502, 'K', 17504-17524, 'K', 17526-17546, 'K', 17548-17568, 'K', 17570-17590, 'K', 17592-17612, 'K', 17614-17634, 'K', 17636-17656, 'K', 17658-17678, 'K', 17680-17700, 'K', 17702-17722, 'K', 17724-17744, 'K', 17746-17766, 'K', 17768-17788, 'K', 17790-17810, 'K', 17812-17832, 'K', 17834-17854, 'K', 17856-17876, 'K', 17878-17898, 'K', 17900-17920, 'K', 17922-17942, 'K', 17944-17964, 'K', 17966-17986, 'K', 17988-18008, 'K', 18010-18030, 'K', 18032-18052, 'K', 18054-18074, 'K', 18076-18096, 'K', 18098-18118, 'K', 18120-18140, 'K', 18142-18162, 'K', 18164-18184, 'K', 18186-18206, 'K', 18208-18228, 'K', 18230-18250, 'K', 18252-1

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:32:31 ; Search time 18 seconds
(without alignments)
4190.602 Million cell updates/sec

Title: US-09-888-615-73
Perfect score: 8563
Sequence: 1 MGAKESRIGFVSYEALRV.....DTSSMDEDFSDYKVKYVLQ 1604

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4531	52.9	1089	1	UBP6 HUMAN
2	1087.5	12.7	981	1	P35125 mus musculus
3	1087	12.7	981	1	UBP6 MOUSE
4	1086	12.7	963	1	UBP4 HUMAN
5	1085.5	12.7	962	1	UBP4 MOUSE
6	903.5	10.6	690	1	UBP6 HUMAN
7	902.5	10.5	699	1	UBP6 MOUSE
8	787.5	9.2	979	1	UBPC SCHPO
9	738.5	8.6	1254	1	UBPC YEAST
10	693.5	8.1	849	1	UBP1 SCHPO
11	665	7.8	1371	1	UBP7 HUMAN
12	498	5.8	1118	1	UBP8 HUMAN
13	446	5.2	357	1	UBP2 CHICK
14	420	4.9	605	1	UBP2 HUMAN
15	412	4.8	353	1	UBP2 MOUSE
16	371	4.3	926	1	UBP4 YEAST
17	343.5	4.0	913	1	UBP4 HUMAN
18	330	3.9	805	1	UBP5 YEAST
19	325.5	3.8	566	1	UBP1 MOUSE
20	321	3.7	565	1	UBP1 HUMAN
21	303	3.5	438	1	UBP4 SCHPO
22	294	3.4	823	1	UBPG HUMAN
23	285	3.3	520	1	UBP3 MOUSE
24	271	3.2	593	1	UBPM HUMAN
25	259	3.0	445	1	UBPB CANFA
26	249.5	2.9	1071	1	UBP7 YEAST
27	245.5	2.9	521	1	UBP3 HUMAN
28	211.5	2.5	717	1	UBPB YEAST
29	210.5	2.5	913	1	UBP0 HUMAN
30	206	2.4	898	1	UBPE DROME
31	203.5	2.4	922	1	UBP7 HUMAN
32	200	2.3	526	1	UBPW MOUSE
33	190	2.2	754	1	UBP9 YEAST

RESULT 1

ID	UBP6 HUMAN	STANDARD;	PRT;	1089 AA.
AC	P35125;			
DC	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Ubiquitin carboxyl-terminal hydrolase 6 (EC 3.1.2.15) (Ubiquitin thioesterase 6) (Ubiquitin-specific processing protease 6)			
DE	(deubiquitinating enzyme 6) (Proto-oncogene TRE-2).			
GN	USP6 OR TRE2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92228503; PubMed=1565468;			
RA	Nakamura T., Hillova J., Mariage-Samson R., Onno M., Huebner K., Cannizzaro L.A., Boghosian-Seil L., Croce C.M., Hill M.;			
RA	"A novel transcriptional unit of the tre oncogene widely expressed in human cancer cells."			
RT	human cancer cells."			
RL	Oncogene 7:733-741(1992).			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=94067315; PubMed=8247125;			
RA	Papa F.R., Hochstrasser M.;			
RT	"The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene."			
RL	Nature 366:313-319(1993).			
CC	-1- FUNCTION: HAS AN ATP-INDEPENDENT ISOPEPTIDASE ACTIVITY, CLEAVING AT THE CARBOXYL TERMINUS OF THE UBIQUITIN MOIETY.			
CC	-1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thioester + H(2)O = ubiquitin + a thiol.			
CC	-1- DISEASE: INVOLVED IN TUMORIGENESIS.			
CC	-1- SIMILARITY: Belongs to peptidase family C19.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL; X63547; CAA45111.1; ..			
DR	EMBL; X63546; CAA45109.1; ALT_INIT.			
DR	PIR; S57868; S22158.			
DR	MEROPS; C19.009; ..			
DR	Genew; HGNC:12629; USP6.			
DR	MIM; 604334; ..			
DR	GO; GO:0004212; F-lysosomal cysteine-type endopeptidase; TAS.			
DR	GO; GO:0003676; F-nucleic acid binding activity; TAS.			
DR	GO; GO:0070748; P-oncogenesis; TAS.			
DR	InterPro; IPR001394; UCH-2.			
DR	Pfam; PF00443; UCH; 1.			

ALIGNMENTS

34	186	2.2	869	1	UBPT MOUSE
35	184.5	2.2	875	1	UBP7 SCHPO
36	183	2.1	785	1	UBP1 HUMAN
37	173.5	2.0	189	1	NCS1 XENLA
38	172.5	2.0	189	1	NCS1 HUMAN
39	172.5	2.0	835	1	UBPQ MOUSE
40	169	2.0	355	1	UBPC HUMAN
41	167	2.0	189	1	NCAH DROME
42	167	2.0	189	1	NCS1 SCHPO
43	166	1.9	1077	1	UBPS HUMAN
44	165	1.9	803	1	UBPE YEAST
45	164	1.9	198	1	GCA1 CHICK

O9es63	mus musculus
O3p785	schizosacch
O94782	homo sapien
O91614	xenopus lae
P36610	homo sapien
O99mx1	mus musculus
O75317	homo sapien
P42325	drosophila
O09711	schizosacch
O96ru2	homo sapien
P38237	saccharomyc
P79880	gallus gall

DR PROSITE; PS00972; UCH 2.1; 1.
 DR PROSITE; PS00973; UCH 2.2; 1.
 DR PROSITE; PS0235; UCH 2.3; 1.
 KW Proto-oncogene; Ubl conjugation pathway; Hydrolase; Thiol protease;
 KW Multigene family.
 FT ACT_SITE 438 BY SIMILARITY.
 FT ACT_SITE 1003 BY SIMILARITY.
 FT ACT_SITE 1011 BY SIMILARITY.
 SQ SEQUENCE 1089 AA; 121883 MW; 6228BCFBED5B980 CRC64;
 Query Match 52.9%; Score 4531; DB 1; Length 1089;
 Best Local Similarity 93.7%; Pred. No. 1.1e-278;
 Matches 851; Conservative 22; Mismatches 31; Indels 4; Gaps 1;
 QY 701 EVNKDMSPEEMSFANSKIDRHVPTEKATGLSLNLTGTCFVNSISQCVSNTPQTQ 760
 DB 182 EVNKDMSPEEMSFANSKIDRQVPTKATGLSLNLTGTCFVNSISQCVSNTPQTQ 241
 QY 761 YFISGHLYELARTNPIGHKGMKACYGDLVQELWSGTOKNVAPLKLRTWIAKYAPRTNG 820
 DB 242 YFISGHLYELARTNPIGHKGMKACYGDLVQELWSGTOKNVAPLKLRTWIAKYAPRTNG 301
 QY 821 FQQDSQELLAFLLDGLHEDLNVRVHEKPYVELKSDGRPDWEVAEADNHLRRNRSTVW 880
 DB 302 FQQDSQELLAFLLDGLHEDLNVRVHEKPYVELKSDGRPDWEVAEADNHLRRNRSTVW 361
 QY 881 DLFHGQLRSQVCKTCGHTSVRPDPNPLSLPLMDSYMHLITVTKLDGTPVRYGLRL 940
 DB 362 DLFHGQLRSQVCKTCGHTSVRPDPNPLSLPLMDSYMHLITVTKLDGTPVRYGLRL 421
 QY 941 NWDEKVTGLKKQLSDLCLNSQIILAEVHGSNIKNFQDNQKVLPSVSGFLCAFEIIVP 1000
 DB 422 NWDEKVTGLKKQLSDLCLNSQIILAEVHGSNIKNFQDNQKVLPSVSGFLCAFEIIVP 481
 QY 1001 VPSIASSTQTFDSSPSTNEMFTLTNGDLPRIPIFNGMPTVVPVCGTEKQFTNGMV 1060
 DB 482 VPSIASSTQTFDSSPSTNEMFTLTNGDLPRIPIFNGMPTVVPVCGTEKQFTNGMV 541
 QY 1061 NGHMLPSDPTGTVIIAHRKQWTELYFLSSQNRSLGMPILVICTVHTKRLDLYD 1120
 DB 542 NGHMLPSDPTGTVIIAHRKQWTELYFLSSQNRSLGMPILVICTVHTKRLDLYD 601
 QY 1121 AVMIQVSLASLPQEAASNAHQDSDSGYQPTFLRWQKDGNSCAWCFWFCRGCK 1180
 DB 602 AVMIQVSLASLPQEAASNAHQDSDSGYQPTFLRWQKDGNSCAWCFWFCRGCK 661
 QY 1181 IDCGRDRAFINAYTAVDWDPTALHLRYQTQSERVVDHESVQSRRAQAPINLDSCLR 1240
 DB 662 IDCGRDRAFINAYTAVDWDPTALHLRYQTQSERVVDHESVQSRRAQAPINLDSCLR 721
 QY 1241 AFTSEELGENEMYCSKCKTHCLATKLDLWRLPPLIILHKKRQFVNGWIKSQIVK 1300
 DB 722 AFTSEELGENEMYCSKCKTHCLATKLDLWRLPPLIILHKKRQFVNDQWIKSQIVK 781
 QY 1301 FRESFDSAFVLPDPALCOHKPTUPQGBELSPRILAREVKKYVDAQSSAGEEDVLLSK 1360
 DB 782 FRESFDSAFVLPDPALCOHKPTUPQGBELSPRILAREVKKYVDAQSSAGEEDVLLSK 841
 QY 1361 SPSSLSANISSPKGSPSSRSKSGTSCPSKNSPNSPRLTGLSKGLRLPQIGSKVKL 1420
 DB 842 SPSSLSANISSPKGSPSSRSKSGTSCPSKNSPNSPRLTGLSKGLRLPQIGSKVKL 901
 QY 1421 SSSKNLDAKENGAGQICELADALSRGHVGGSPQLVTPQDHEVALANGFLYEACG 1480
 DB 902 SSSKNLDAKENGAGQICELADALSRGHVGGSPQLVTPQDHEVALANGFLYEACG 961
 QY 1481 N---GYSGNGLGNHSESDTDQREDTRIKPIYLNLYAISCHSILGCHGVYVYAKNPNC 1536
 DB 962 NCGGNGNGQGNHSESDTDQREDTRIKPIYLNLYAISCHSILGCHGVYVYAKNPNC 1021
 QY 1537 KWCYNDSCKELHPDEIDTDSAYILFYEQGIDYAPLPKTDGKQWADTSSDDEPESD 1596
 DB 1022 KWCYNDSCKELHPDEIDTDSAYILFYEQGIDYAPLPKTDGKQWADTSSDDEPESD 1081

QY 1597 YKCYVULQ 1604
 DB 1082 YEKYSMLQ 1089
 RESULT 2
 UBPF MOUSE
 ID_UBPF_MOUSE STANDARD; PRT; 991 AA.
 AC Q8R5H1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 15 (EC 3.1.2.15) (Ubiquitin
 thiolesterase 15) (Ubiquitin-specific processing protease 15)
 DE (Deubiquitinating enzyme 15).
 GN USP15.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
 RC STRAIN=C57BL/6;
 RX MEDLINE=22419901; PubMed=12532266;
 RA Angelats C., Wang X.-W., Jermin L.S., Copeland N.G., Jenkins N.A.,
 Baker R.T.;
 RT "Isolation and characterization of the mouse ubiquitin-specific
 protease Usp15.";
 RL Mamm. Genome 14:31-46(2003).
 [2]
 SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina, Testis, and Thymus;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Experimental confirmation may be lacking for some
 isoforms;
 CC Name=1;
 CC IsoId=Q8R5H1-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8R5H1-2; Sequence=VSP_005262; VSP_005263;
 CC Name=3;
 CC IsoId=Q8R5H1-3; Sequence=VSP_005264;
 CC Name=4;
 CC IsoId=Q8R5H1-4; Sequence=VSP_005265; VSP_005266;
 CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in
 testis, heart and liver.
 CC -1- SIMILARITY: belongs to peptidase family C19.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1] _TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain Cortex;
 RX MEDLINE=95303480; PubMed=7784062;
 RA Gray D.A., Inazawa J., Gupta K., Wong A., Ueda R., Takahashi T.;
 RT "Elevated expression of Unph, a proto-oncogene at 3p21.3, in human
 lung tumors.";
 RL Oncogene 10:2179-2183(1995).
 RN [2]
 RP REVISIONS.
 RA Gray D.A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=98124180; PubMed=9464533;
 RA Frederick A., Rolfe M., Chiu M.I.;
 RT "The human UNP locus at 3p21.31 encodes two tissue-selective,
 cytoplasmic isoforms with deubiquitinating activity that have reduced
 expression in small cell lung carcinoma cell lines.";
 RL Oncogene 16:153-165(1998).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 ubiquitin + a thiol.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=UNPCL;
 CC IsoId=Q13107-1; Sequence=Displayed;
 CC Name=UNPES;
 CC IsoId=Q13107-2; Sequence=VSP_005258;
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U20657; AAB72237.1; -;
 DR EMBL; AF017305; AAC27355.1; -;
 DR EMBL; AF017306; AAC27356.1; -;
 DR PIR; T09478; T09478.
 DR MEROPS; C19.010; -;
 DR Genew; HGNC:12627; USP4.
 DR MIM; 603486; -;
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0004212; F:lysosomal cysteine-type endopeptidase; TAS.
 DR GO; GO:0004843; F:ubiquitin-specific protease activity; TAS.
 DR InterPro; IPR006615; DUSP.
 DR InterPro; IPR001394; UCH-2.
 DR Pfam; PF00443; UCH; 1.
 DR SMART; SM00695; DUSP; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS50235; UCH_2_3; 1.
 KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family;
 KW Proto-oncogene; Alternative splicing.
 FT ACT_SITE 311 311
 FT ACT_SITE 873 873
 FT ACT_SITE 881 881
 FT VARSPIC 232 279
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT KSSTAPSRNFTTSPKSSASPYSSVSSASLIANGDSTSTCGMH
 FT SSGVSRG -> N (in isoform UNPES).
 FT /ftid=VSP_005258.
 FT MUTAGEN 311 311
 FT CONFLICT 373 373
 FT CONFLICT 744 744
 FT CONFLICT 744 744
 FT SEQUENCE 963 AA; 109564 MW; 1B62B752F9410CD7 CRC64;

Query Match 12.7%; Score 1087; DB 1; Length 963;
 Best Local Similarity 23.9%; Pred. No. 8.5e-61;
 Matches 327; Conservative 168; Mismatches 313; Indels 560; Gaps 40;
 QY 367 RPATPEEGQIIIRGWLERSRYGLQAGHNWFTISMQWQKQKYYKYDANPVVIBPSSVL 426
 DB 10 RPAETQKSEL--GPLMRTT---LQGAQWYLDSRWFKWKQKYYGFD----- 52
 QY 427 NGGKYSFGTAAPHPMEQVEDRIGSSLSYNTTTEKSDNISTASEASETAGSGFLYSATPG 486
 DB 53 SWDMYNG----- 60
 QY 487 ADVCFARCHNTSDNNQCILGANGNILLHLNPKQPGADINQPLVTOEPVKATSLTEGGR 546
 DB 61 -----EHLNF-----PGIDNSGLFS-DP-----ESQT 82
 QY 547 LKRTQLIHGRDIYEMVPEVPVWALYHWTGANIALPRPVKNSKTDIPELEFPFR----- 600
 DB 83 LKE--HLIDELDYLVPTTEAMNKLNNWYGC-VEGQOPIVRK---VVEHGLFVVKHCKVEV 135
 QY 601 YLLFLRQOPATRTQCSNIWNNGNVPSNAPLKVLAVTGC-FSRMQTIKEIHEVLSQRL 659
 DB 136 YLELKL-----CENSDPTNVL--CHFSKADIIATIEKEMKDLF 173
 QY 660 RI-KBEDMLNLYNSENLYTLDDHDKLEYLKIODEQHLVIEVRNKMDSHPEE----- 712
 DB 174 NIPARETRLMNKMYSNTYEQLSKLDNTVQDAGLYQGQVLVEPQNEGTWPRQTLQSKS 233
 QY 713 -----NSFIAN-----SSKIDR-----HKVP 728
 DB 234 STAPSRNFTTSPKSSASPYSSVSSASLIANGDSTSTCGMHSSGVSRRGGSGFASVNCQPPP 293
 QY 729 T---BKGGATGLSNLGNCTCFMNSSIQCVSNTQPLTOYFISGRHLYELNRTNPIGKGMHAK 785
 DB 294 SSHIQGL-CGLNGLNCTCFMNSALQCLNTAPLTDYFLKDEVEAEINRDNPLGKKGTAIE 353
 QY 786 CYGDLVQELWSTQKNVAPLKLRTWIAKYARFNGFOQDSQELLAFLLDGLHEDLNRVH 845
 DB 354 AYAEILKQMDVADAHVAPMFKTOVGFPAPQFGYQOQDSQELLAFLLDGLHEDLNRVK 413
 QY 846 EKPYYVELKDSQGRPDWEAAEAWDNHLNRNRSIVVDLPHGQLRSQVCKCTCGHISVRDPP 905
 DB 414 KKPYLELDKANGRPDAVAKAENWENHRLNDSIVDTTHGLFKSLVCPKAKSVITDPP 473
 QY 906 FNFLSLPLPMSYMELEITVTKLD-GTTPVRYGLRLNMDKEYTGLKKQLSDCLGNSQI 964
 DB 474 FCYLTPLPLKKDVRMEVFLVPADPHCRPTQYRVTVPLMGAVSDICEALSRLSGTAENM 533
 QY 965 LLAEVHGSNI-KNFPQDNQKVELSVSGFLCAFEIPVPVSPISASSPTQDTFSSSPSTNEM 1023
 DB 534 VVADVYNHRFKIFQWDEGLNHIHPRDDIFVEV-----CSTSVGGSEC 577
 QY 1024 FTLTNGDLPRPIFIPNGMNPNTVWPCGTEKNTNGMNGHMPSLPDSPTGYIIIVHHRKM 1083
 DB 578 VTL-----PV----- 582
 QY 1084 METELYFLSSQKNRPS-----LFGMLPIVPCVTH-TRKDLVDYDAWIVQVSRLL----- 1129
 DB 583 -----YFRERKSRPSSTSSALYQGPLLVSVPKHKLTLSLYQAVCDRISRYVYKQPLP 636
 QY 1130 ----ASPLPPOEASNAQDC----DDSMGYQVPFTLRVVQKDGNSCAWCPWYRCRCCKI 1181
 DB 637 DEFGSSPLEPGACNGSRNSCEGEDEEMEHQEGKEQLSETEGS----- 680
 QY 1182 DQGED-----RAF-----IGNAY-----I 1195
 DB 681 --GEDEPGNDPSETTKKIKQKQPCPKRLFTESLNSYNGTADINSIAADGKLLKLNSTL 738
 QY 1196 AYDWDPTALHLYQTSQBRVDEHESVEQSRRAOAEPINLDSCLRAFTSEELGENEMY 1255
 DB 739 ANDWDRETRLYYDEQSEAEYKHKVSMLOPKQKKKTVALRDCIELFTTMTLEHDPWY 798
 QY 1256 CSKCKTHCLATKGLDLWLPILIIHLKRFOFVNGRW-IKSKIVKFPRESFDSAFILVP 1314

```
Db 799 CPNCKGHOATKFDLWLSPLKILVHLKRFY-NRYWRKLDTVVEFFIRGLNMFEP--- 854
QY 1315 RDPALCOHKLPTQGDSELPRLAREVKVDAQSSAGEEDVLLSKSPSSIANIISPK 1374
Db 855 ----- 854
QY 1375 GSPSSSRKSGTSCPSKSNSSPNSPRTLGRSKRLRLPQIGSKNKLSSKENLDASKENG 1434
Db 855 ----- 854
QY 1435 AQOICELADALSRGHVLSQSQBELVTPQDHEVALANGFLYEHACNGYNGQJGNHSEE 1494
Db 855 -----VGNLS----- 859
QY 1495 DSTDDQREDTRIKP-IYNLYALSCHSGILGGHYVYAKNP-NCKWYCVNDSSCKELHPD 1552
Db 860 -----ARPYVDLIAVSNHYGANGVGHYAYAKNKLNGWYFYDSDNVLASED 908
QY 1553 EIDTDSAYILFYEQGIDYAOFLPTDQKQVAD-----TSSMDEDFSD 1596
Db 909 QIVTKAAVYLFYQRDDDFYK-TPSLSSSGSDGGRTPSPSSQGGFGDD 955

RESULT 4
UBP4_MOUSE
ID UBPA_MOUSE STANDARD; PRT; 962 AA.
AC P35123; OS4704;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 4 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 4) Ubiquitin-specific processing protease 4)
DE (Deubiquitinating enzyme 4) (Ubiquitous nuclear protein).
GN USP4 OR UNP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98267137; PubMed=9602026;
RX di Fruscio M., Gilchrist C.A., Baker R.T., Gray D.A.;
RA "Genomic structure of Unp, a murine gene encoding a ubiquitin-specific
RT protease";
RL Biochim. Biophys. Acta 1398:9-17(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Di Fruscio M., Gilchrist C.A., Baker R.T., Gray D.A.;
RT "Genomic structure of Unp, a murine gene encoding a ubiquitin-specific
RT protease";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 169-962 FROM N.A.
RX MEDLINE=93330575; PubMed=8336951;
RA Gupta K., Copeland N.G., Gilbert D.J., Jenkins N.A., Gray D.A.;
RT "Unp, a mouse gene related to the tre oncogene.";
RL Oncogene 8:2307-2310(1993).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: OVEREXPRESSION LEADS TO ONCOGENIC
CC TRANSFORMATION OF NIH 3T3 CELLS
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
or send an email to license@isb-sib.ch.
-----
CC EMBL; L00681; AAB82339.1; --
CC DR EMBL; AF026469; AAC53587.1; --
CC DR MEROPS; C19.010; --
CC DR MGI; MGI:98905; Usp4.
CC DR InterPro; IPR006615; DUSP.
CC DR InterPro; IPR001394; UCH-2.
CC DR Pfam; PF00443; UCH_1.
CC DR SMART; SM00895; DUSP; 1.
CC DR PROSITE; PS00972; UCH_2_1; 1.
CC DR PROSITE; PS00973; UCH_2_2; 1.
CC DR PROSITE; PS0235; UCH_2_3; 1.
CC DR PROSITE; PS0235; UCH_2_3; 1.
CC KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family;
CC FT ACT_SITE 311 311 BY SIMILARITY.
CC FT ACT_SITE 872 872 BY SIMILARITY.
CC FT ACT_SITE 880 880 BY SIMILARITY.
CC FT DOMAIN 765 772 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT CONFLICT 292 292 A -> P (IN REF. 2).
CC SQ SEQUENCE 962 AA; 108280 MW; 9553CAD9B4857360 CRC64;

Query Match 12.7%; Score 1086; DB 1; Length 962;
Best Local Similarity 24.2%; Pred. No. 9.8e-61;
Matches 322; Conservative 164; Mismatches 322; Indels 522; Gaps 38;

QY 390 LQAGHNWFIISMOWQWQWKEYKYKIDANPVVIBPSSVLNGKYSFGTAHPMQVEDRIGS 449
Db 28 LQGAQWYLDSEWFKWKYVGF-----SWDMYNG----- 60
QY 450 SLSYVNTTEKFSNDISTASEASETAGSGLYSATGADVCFARQHTSDNNNQCLLGAN 509
Db 61 -----EHNLF----- 65
QY 510 GNILLHLNPKGAINQPLVTPQEPVKATSLTEGRLKRTPLQIHGRDYEMVPEPWEA 569
Db 66 -----POPIDNSGLFS-DP-----ESQTLKE--HLIDELDYLVPAEAMNK 103
QY 570 LYHWYGANALPRPVKNSKTDIPELELPPR-----YLLFLRQOPATRTQGSNIWVNG 623
Db 104 LLNMYGC-VEGQOPIVRKVVD-----GLFVKHKVEVILLEKL----- 142
QY 624 NVPSNPAPLKRVLAYTGC-FSRMQIKETHEVLSORLRI-KEEDMLRLWYNSENVLTLD 681
Db 143 ---CENSDDTNVLS---CHFSKADTIATIEKMRKLFNIPAEERTRLMNKYMSNTYEQLS 196
QY 682 DEDHKLEYLKIODEQHLVIEVRNKMSPPE----- 712
Db 197 KLDNTIQDAGLYQGVVIEPQNEEDGTWPQSLQSKSSTAPSRNFTTSKPSASPYCSVS 256
QY 713 MGFIAN-----SSKIDR-----HKVPT---EKATGLSNLGNTCFMNSS 748
Db 257 ASLIANGDSTNSGMSHSGVSRGSGFSAVNCQBPSPHIQPGLCGLNGLNCTCFMNSA 316
QY 749 IOCVSNTQPLTOYFTSGRHLVELNRTNPIGMKGMKACVGDIVQELWSGTQKQVAPLKL 808
Db 317 LQCLSNAPLTFYFLKDEYAEINRDNFLMGKGEAEAYAEILKOMWGRDTHVAPRMFK 376
QY 809 WTIAKYAPRNFNGFOQDSQELLAFLLDGLHEDLNVRHEKPYVELKSDGRPDWEAAEAW 868
Db 377 TQVGRFAPQFSGYQOQDSQELLAFLLDGLHEDLNVRKPKVLEPKDANGRPDAVVAKEAW 436
QY 869 DNEHLRNSIVVDLPHGQLRSQVKCTCGHISVRPDPNPLSLPLPMDSYHLEITVKL 928
Db 437 ENHRLRNSIVVDLPHGFLFKSLTVQPECAKSVTPDPFCYLTLPPLKDKRIMEVFLVPA 496
QY 929 D-GTTPVRYGLRNLNDEKTYTGLKKQLSDLCGLNSQIILAEVHGSNI-KNFPQDNQKRL 986
Db 497 DPQCRPIQRYVTPLMGALSDICEALSGLSGIAEAMVVTVDYVNRPHKIQMDEGLSHI 556
QY 987 SVSGFLCAFEIPVPVSPISASSPTQDFSSSPSTNEMTTLTNGDLPRPIFINGMPNTV 1046
Db 557 TPRDDIFVYEV-----CNTSDMGSECITL-----PVYF----- 584
```


RA Whitehead S.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBSJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stables M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL; U44839; AAC50450.1; -;
 DR EMBL; AL096791; CAD20056.1; -;
 DR EMBL; BC000350; AAH00350.1; -;
 DR MEROPS; C19.014; -;
 DR Genew; HGNC:12609; USP11.
 DR MIM; 300050; -;
 DR GO; GO:0004197; P: cysteine-type endopeptidase activity; TAS.
 DR GO; GO:0004843; F: ubiquitin-specific protease activity; TAS.
 DR InterPro; IPR001394; UCH-2.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH 2.1; 1.
 DR PROSITE; PS00973; UCH 2.2; 1.
 DR PROSITE; PS0235; UCH 2.3; 1.
 DR Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 53 53 BY SIMILARITY.
 FT ACT_SITE 607 607 BY SIMILARITY.
 FT ACT_SITE 615 615 BY SIMILARITY.
 FT CONFLICT 216 216 P -> L (IN REF. 3).
 SQ SEQUENCE 690 AA; 79128 MW; A7DDE9BD3ADE259 CRC64;
 Query Match 10.6%; Score 903.5; DB 1; Length 690;
 Best Local Similarity 25.9%; Pred. No. 2.1e-49;
 Matches 249; Conservative 122; Mismatches 231; Indels 359; Gaps 23;
 QY 700 IEVRNKMDSWPREMFIANSKIDRHK-VPTKKGATGLSNLGNCTCFMNSSIQCVSNTPQL 758
 DB 1 METRKGDTGTPSAQLHVNNNNSEDEDFKQPGICGLTGLNCTCFMNSALQCLSNVPL 60
 QY 759 TQYFTSGRHLNELNTNFGMKGHAKYVDGLVQELWSGTQKNVAPLKLRTIAYAPRF 818
 DB 61 TSYLNNCLLELNFNPLMGKTAAYADLVKQAWSGHRISVPHVFNKVGHFASQF 120
 QY 819 NGFQODSOELLAFLLDGLHDLNVRHVKPYVELKDSGRPDWEAAEDNHLRRNRSI 878
 DB 121 LQYQQHDSOELLFLLDGLHDLNVRHVKPYVELKDSGRPDWEAAEDNHLRRNRSI 878

QY 879 VVDLHGLQLRSQVKCTCGHISVRDPENFLSLPLPMSYMELEITVTKLD-GTTPVRYG 937
 DB 181 IVDTHGLFKSLVCPDCGNVSVTFDFPCILYVPLISHKRVLEVFIFPMDPRRPEQHR 240
 QY 938 LRINMDEKTYGLKQLSDLCGLNSQILAEVHGSNIKFNFPQDNQKVLRSVSGFLCAPEI 997
 DB 241 LVVPKKGKISDLVALSKHTGISPERMMVADVFSHR-----FYKLYQL 283
 QY 998 PVPVSPISASSPTQDTDFSSSPSTNEMFILTNG-----DLPRPIFINGMNTVV 1047
 DB 284 EEPLSIL-----DRDIFVYVSGRIEAGSREDIVVPYLRERTP----- 326
 QY 1048 PGTETKNTNG-----MVNGH--MPSLPDPSPTFGYIIAVERKMRTELYFLSSQKNRPSLF 1101
 DB 327 ----ARDYNSYGLMLFGLHLLSVPRDRFTW----- 355
 QY 1102 GMPDIVPCTVTRKCDLYDAVWIVQR----- 1128
 DB 356 -----EGLNVLMYRLSRYVTKNSDEDDGDEKDEEDKDDVPGPSTGG 401
 QY 1129 -LASPLPQO--EASNHAQDC---DSSMG-YOYP-----FTLRVVKDCNSCAWCPW 1172
 DB 402 SLRDPPEQAGPSSGVTRNCPFLDNCIGTSONPPRRRKQLFTLQTVNSNGTS----- 455
 QY 1173 YRFRCKIDCGEDRA-----FIGNAYIAVDWDPALHLRYQTSQERVVDEHESVEQRR 1227
 DB 456 -----DRTPSEVHAQFYAIDWPEMKRYVDEVEAGYVHKDCVGYMK 502
 QY 1228 AQAEPIINDSLRAFTSEELGENEMYCKCKTHCLATKLDLWLPILLIHLKRFQF 1287
 DB 503 --KAPVRLQECLELFTTETLEKENFTWPCSCQKQLATKLDLWLPILLIHLKRFQF 560
 QY 1288 VNGRWTKSKIVKFPRESFDPSPAFVPRDPALCOHKPLTPQGDSELPRLAREVKKYDA 1347
 DB 561 TKFSREKDLTVFPPRDLDFEFVI-----QPQNE----- 591
 QY 1348 QSSAGEEDVLLSKSPSSLSANISSPKSPSSRSKSGTSCPSKNSNSPRTLGRSGK 1407
 DB 592 ----- 591
 QY 1408 RLRLPQIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLSQSPQLVTPQDHEVA 1467
 DB 592 -----SNPE----- 595
 QY 1468 LANGFLYEHAACGVNSQNGQLGNHSEEDSTDQREDTRIKPIYNDLYATSCHSIGLGGHY 1527
 DB 596 -----LYK-----YDLIAYSNHYGMRDGHY 616
 QY 1528 VTYAKNPNC-KWVCYNDSSCKELHPDEITDTSAYILFYEQQIGIDYAQFLPKTDGKKMADT 1586
 DB 617 TTFACNKGSGWHYFDNDSVSPVNEQIESKAAVLYFYQRQ--DVARRLLSPAGSSGAPA 674
 QY 1587 S 1587
 DB 675 S 675
 RESULT 7
 UBPE_MOUSE
 ID UBPE_MOUSE STANDARD; PRT; 699 AA.
 AC Q99K46;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 11 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 11) (Ubiquitin-specific processing protease 11)
 DE (Deubiquitinating enzyme 11).
 GN USP11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;

```

RN SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC005470; AA05470.1;
DR MEROPS; C19.036;
DR MGD; MGI:2384312; Usp11.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS02035; UCH_2_3; 1.
DR KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
DR ACT_SITE 52 52 BY SIMILARITY.
DR ACT_SITE 617 617 BY SIMILARITY.
DR ACT_SITE 625 625 BY SIMILARITY.
DR FT SEQUENCE 699 AA; 80145 MW; AB86D90579AE7C98 CRC64;
SQ

```

```

Query Match 10.5%; Score 902.5; DB 1; Length 699;
Best Local Similarity 25.6%; Pred. No. 2.4e-49;
Matches 249; Conservative 123; Mismatches 227; Indels 373; Gaps 24;

QY 700 IEVTKNMSWPEMSFTANSKIDRHVKTEKGNATGSLNIGTCFNMSSIQCVSNTPQLT 759
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1 METAKDGTWPSAQLCGMNNIPDEDEDFQGPQGTGTLNGLCFMNSALQCLSNVPQLT 60
QY QYFTSGHLYELNRTNPIGKGMHAKCGVLQVQLWSGTOKVWAPLKLRTIAKYAPREN 819
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 61 EYFLNNRYELNRTNPIGKGMHAKCGVLQVQLWSGTOKVWAPLKLRTIAKYAPREN 120
QY 820 GFQOQDSQELLAFLLDGLHEDLNKRVHEKPYVELKDSGRPDWEAAEADNHLRRNSIV 879
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 121 GYQHQDSQELLFLDGLHEDLNKRVHEKPYVELKDSGRPDWEAAEADNHLRRNSIV 180
QY 880 VDLPHGQLRQVCKTCTGHSVDPNPNFLSLPLPMDSYNHLITIKLDTGTPVRYGLR 939
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 181 VDTFHGLFKTLVPCDGNVSTFDPPCYLSVLPVCSREVLEVPFVPMVPMVPMVPMV 234
QY 940 LNMDEKVTGKQKQLSDLC-----GLNSQKILLAEVHGSNNKNNFQDNQKVELSVGFL 992
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 235 PEQRVVVPPKGNISDLCTVALSTHTSVAPDKMIVADVFSHRFKYLIQ-----LEDPLSGIL 290

```

```

QY 993 -----CAFEIPVPVSPISASSPTQTDFFSSSPSTNEMFTLTNGDLPRPIFNGMPTVV 1047
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 291 DRDDIFVIVTGRIEVEGS-----RDDIVVPVILERTIPS----- 326
QY 1048 PCETERKNTNGMVGHPMLSPDPTGYIIAHRKMMRTLEYFLSSQKNRPSLFGMPLIV 1107
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 327 -----RDYNS-----YGLI-----LFGHPLIV 345
QY 1108 PCTVHTRK-----KDLVDVAVIOVSR-LASPLPQEASNAHQDCDD----- 1147
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 346 --SVPRDRFSWGLYNILMRUSRYVTRKTSDEDDGDEKVEDDEDVDDSSSEKEE 403
QY 1148 -----SMGY-----QYP-----FTLVVQKDG 1164
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 404 MSAPTVDNDGTREAEQEQAGTSQVTERCPSLLDNLRSQWPPRRRRKQLFTLQTVNSG 463
QY 1165 NSCAMPWTRFCGCKIDGCDRA-----FIGNAIVDWDPTALHLRYQTSQERVVDH 1219
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 464 TS-----DRTTSPEEMQTPYIAMDWEPMKRRYYDEVEAEGYVKH 504
QY 1220 ESVEQSRRAQABPINLDSCLRAFTSEELGENEMYCKKTHCLATKCLDLWRLPPLI 1279
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 505 DCV--GYMLKKSPVOLKECIKFTTVELEKENPWYCSSCKQHQLATKCLDLWMLPEVLI 562
QY 1280 IHLKRPQFVNGRWIKSQKIVKPPRSFDPSPAPLVRDPALCOHKPLTPQGDLSBPRLA 1339
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 563 IHLKRFSSKIREKLDLTIVQPIRDLPSEFVI-----KP----- 598
QY 1340 REVKKVDAQSAGEEDVLLSKSPSSLSANISSPKGSPSSRSKSTSCFSSKNSSPNSSP 1399
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 599 -----KNESPD----- 605
QY 1400 RTLGRSKGLRLPQIGSKNKLSSKENLDASKENGAGQICELADALSGRHVLGSGQPELV 1459
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 606 ----- 605
QY 1460 TPQDHEVALANGFLYEACNGYNGQGNHSEEDSTDQREDTRIKPIYNLYAISCHS 1519
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 606 -----LYK-----YDLIAVSNHY 618
QY 1520 GILGGGHVTVAKNENC-KWYCYNDSSCKELHPIDTDSAYILFYEQQGDIDYAOFLPKT 1578
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 619 GCMRGDGHYTFACNKDSQWHYLDNDSVFNENQIESKAAVLYFYQKQDV----- 669
QY 1579 DGKKWADTSSMD 1590
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 670 -GRQSQTSSTD 680

```

```

RESULT 8
UBPC_SCHPO STANDARD; PRT; 979 AA.
ID UBPC_SCHPO
AC O60079;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable ubiquitin carboxyl-terminal hydrolase 12 (EC 3.1.2.12)
DE (Ubiquitin thiolesterase 12) (ubiquitin-specific processing protease
DE 12) (Deubiquitinating enzyme 12).
GN UB12 OR SPCC1494.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Art R., Robben J., Grymonprez B., Weltjens I., Vanstreels E., Kieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G., Dega R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Shpakovski G.V., Ussery D., Barrell B.G., Nurse P., Potashkin J., "The genome sequence of *Schizosaccharomyces pombe*."; Nature 415:871-880(2002).

CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O = ubiquitin + a thiol.

CC -!- SIMILARITY: Belongs to peptidase family C19.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL; AL023776; CAAL19303.1; -

DR PIR; T41006; T41006

DR GeneDB; Spombe; SPCC1494.05c; -

DR InterPro; IPR006615; DUSP.

DR InterPro; IPR001394; UCH-2.

DR Pfam; PF00443; UCH; 1.

DR SMART; SM00695; DUSP; 1.

DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS00973; UCH_2_2; 1.

DR PROSITE; PS0235; UCH_3; 1.

DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.

KW ACT SITE 320

FT ACT SITE 320

FT ACT SITE 927

FT ACT SITE 935

FT ACT SITE 935

SQ SEQUENCE 979 AA; 111967 MW; A705444BCDB437D2 CRC64;

Query Match 9.2%; Score 787.5; DB 1; Length 979;

Best Local Similarity 22.9%; Pred. No. 7.8e-42;

Matches 283; Conservative 162; Mismatches 414; Indels 377; Gaps 37;

QY 432 SFGTAAPMEQVEDRIGSSISVYNTTEKFSNDISTASEATAGSGFLYSATPGADVCF 491

Db 19 SLSESSQSSNNWDISQKSLGDAE--ISKNLPSIAEKQKQLIG----- 61

QY 492 ARQNTSDNNQCILLGANGNILLHLNPK-----PGAINQPLVTOEPVKATS 539

Db 62 ----ELVNNQPELGGVDNYILSYWYERLCYLAEDGPPGPDQEDIA----- 108

QY 540 LTLEGGRLKRTPLQIHGRDYMVEPPVRYALHYWG-ANLALRPVTKNSKTDIPEL--E 596

Db 109 -DLTGILK--PDLOEIDFIIRSDVLLVRYGLKGPFPFRETNLGSGSHPLHVE 165

QY 597 LFPRLYLFLRQOPATRTQSQNIWNNMGVSPNAPLKRVLAYTGCFSRMOTIKEIHEYL 655

Db 166 VYPPIFSLTLLSTNAVDA NES-----HKPKKISL-----SSKSTLEDLEGV 207

QY 656 SQRARIKEEDNRMLYNSENVLTLDDEHKLKYLKIQDQHLVIEVRNKM----- 707

Db 208 KYTSLSPDSQRLRWVTDQPLHRTIDPS---SPKINSKEIIDFLEKSKTLVELGMDSS 264

QY 708 -----SWPEE-----MSPI-----ANSSKIDRHKVPTEKATGLSNLGNCTCF 744

Db 265 CSLVAECMINETVPVDRALRLQFLIOQRNNSNEBQKQKRVF---GTGSLNLGNCTCY 321

QY 745 MNSSIQCVSNTQPLTOYFISGRHLVELNRTPIGWKGMKACYGDLVQELWSGTQKNVAP 804

Db 322 MNSALQCLTHRELEDFTSDEWKNQVNESPLNGWGQVASIFASLKLKSLSPHSSEFAP 381

QY 805 LKLRWTIAKAPRNGFQQQSQELLAPLLDGLHEDLNVRHEKPYV---ELKDSGDGRDW 861

Db 382 RQKATIGKFNHSLFGYQQDSQSFLAPLLDGLHEDLNRIYKQPYTKPDIYEVEDEKIK 441

QY 862 EYAAEAWDNLRRNRSIVVDLPHGLRSQVCKTCGHSVRFDPDFNFIPLPLPMDSYMHL 921

Db 442 NTAECWRLKRLNDSLIVDLFQGYRSTIVCPVNTVSTITDFDFMDLTLPPLPKQVWSH 501

QY 922 EITVKLD-GTTPVRYGLRLNDE-----KYTLKKQLSDLCGLNSEQILLAEVHGSN 973

Db 502 TVTFIPADTLTPLAIEVVLESKAATIEDLVKYVAEKSGCSDY-----RKILVTETYKGR 556

QY 974 IKNF-PODNQKVRLSVG-----FLCAFEIP-----VPSPISSASSPTQTD-FSSSP 1018

Db 557 FYRFTQLSKSLMEISEDEIYELERPYDEGSDILVPVYHISDDTSNANSYMSR 616

QY 1019 STNEMFTL-TTNGDLPRPIFIPNGMPTVPGTEKQFTN-----GMVNGHMPSLPSPF 1072

Db 617 DFGEHFFVLQSDNEVTDASFISEKLKLYQQFTTLKNLKNIDSLSELGH-----EDEQV 672

QY 1073 TGYIIAVRKMRTELTYLLSSQKVR---PSLFGMPLI-VPCVTHTKKKLDYDAWVQVS 1127

Db 673 QKGPLDVMDSHSQTPLFEMRVFHDRFEKIPITGWMSVNSLPLTERDKD----- 722

QY 1128 RLASPLPQEAENHAQDCDDSMGYQYPTLRVQVKGNSCAWCPWYRCRCCKIDCGS-D 1186

Db 723 -LESTVDFLDASIEEEDDSFKDVAFGSYEPESKSNEN-----TKLTAKEND 769

QY 1187 RAFIGNAYIADWDPTALHLRYQT-----SQERVVDEHESVEQSRAQAP 1232

Db 770 RLLIQGDLVCEWPEKSYQFVSAPSPQMGSRSLWLESKITLSDKKDDSDSR-----T 824

QY 1233 INLDSCLRAFTSEELGENEYCYCKKTHCLATKLDLWLPPIILHLARFQFVNGRW 1292

Db 825 ITLNDCLDFEKTQELGDEPWYCTCKEPRQASQKQMEIMRCPEILIFHLKRFSSERRFR 884

QY 1293 IKSQRIKVFPPRESFDPASFLVRDPALCOHKPLTPQGDSELPRLAREVKVDAQSSAG 1352

Db 885 DKIDDLVEFP----- 894

QY 1353 EDVLLSKSPSLSANIISSPKGSPSSRSKSGTSCPSKNSPNSPRTLGRSGRLRLP 1412

Db 895 ----- 894

QY 1413 QIGSKNKLSSKENLDASKENGAGQICELADALSRGHVGGSPQLVTPQDHEVALANGF 1472

Db 895 -----IDNLDSMTGSKYKLS----- 911

QY 1473 LYEHACNGVYNGOLGNHSEEDSTDQREDTRIKPIYNYLVAISCHSGILGGHYTVYAK 1532

Db 912 -----KENPKL--IYELAVDNYHGGGLGGHYTAFAK 941

QY 1533 NP-NCKWYCYNDSCKELHPDIDTDSAYILFYEQQ 1567

Db 942 NPDNGQFICFDDSRVTPVCPEETWTSAYLLFYRRK 977

RESULT 9

UBPC YEAST ID UBPC YEAST STANDARD; PRT; 1254 AA.

AC P39538;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Ubiquitin carboxyl-terminal hydrolase 12 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 12) (Deubiquitinating enzyme 12).
GN UBPL2 OR YL197W OR J0340.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95274326; PubMed=7754713;
RA Fournelle B., Cosser F., Goffeau A.;
RT "The sequence of a 36 kb segment on the left arm of yeast chromosome
RT x identifies 24 open reading frames including NUC1, PRP21 (Spp91),
RT CDC6, CRY2, the gene for S24, a homologue to the aconitase gene ACO1
RT and two homologues to chromosome III genes.";
RL Yeast 10:1235-1249(1994).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X77688; CAA54762.1; -
DR EMBL: Z49472; CAA89492.1; -
DR PIR: S46636; S46636.
DR MEROPS: C19.0PW; -
DR SGD: S0003733; UBF12.
DR InterPro: IPR006615; DUSP.
DR InterPro: IPR001394; UCH-2.
DR Pfam: PF00443; UCH; 1.
DR SMART: SM00695; DUSP; 1.
DR PROSITE: PS00972; UCH_2_1; 1.
DR PROSITE: PS00973; UCH_2_2; 1.
DR PROSITE: PS0235; UCH_2_3; 1.
KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT_SITE 373 373 BY SIMILARITY.
FT ACT_SITE 1060 1060 BY SIMILARITY.
FT ACT_SITE 1068 1068 BY SIMILARITY.
SQ SEQUENCE 1254 AA; 143191 MW; 349AFA4C4CEE0EAS CRC64;

Query Match 8.6%; Score 738.5; DB 1; Length 1254;
Best Local Similarity 21.8%; Pred. No. 1.5e-38;
Matches 296; Conservative 159; Mismatches 384; Indels 521; Gaps 47;

QY 445 DRIGSLSYVNTTEKFSNDISTAS-----EASSETAGSLYSATPGADVCFARQHTS 498
DB 29 DELGVDL--MNVLDK--DEIKQESVPVSDREIEDTESDASAVSFASANELIAEPHAAS 84
QY 499 DNNNQCLLGANG----NILLHLNPKCAIDNPLVTPQEPVKATSLT-LEGRLKRTPOL 553
DB 85 ETN-----LGTNGQGRNVL-----EQORDVVARLIEENKETQKGGDKVCIIVPKV 129
QY 554 IHG-----RDYE-----MVPEPYWRALYHWYGANL 578
DB 130 WYDKFPDPVTPDPIGPIINTRMICRDPENFVLEDYNRCPVLSIAEPVFNFLSIYGMTS 189
QY 579 ALPRVINKSTDIPELIPRY-----ILFLQQPATRTQQ-----SNTWNGNVPS 627
DB 190 G-SYPTVNLVNTQTGTLETYKWFPRHLHYLTKQDKRRRGQDDSIWLSM----- 243
QY 628 FNAPLKRVLATGCFSPMTIKEIHE-----YLSQRLRKEEDMLW-----LYNS 673
DB 244 -----SALNVLVDLVEKSMNLFPEKADHLVNAVDFKLVFVSEGSDIATD 288
QY 674 ENYLTLLDD--EDHKLEVYIKI-----ODEQHLVIEVR--NKDMSWP 710

DB 289 SNVSTELNSSYEITPLOFLELPKILLIDPMFENRLDKITSPSDLVIEIKIEGHNHP 348
QY 711 EMSFTIANSKIDRHVKPTKEKGATGSLNLTGNTCFMNSSIQCVSNTQPLTQYFISGHLVE 770
DB 349 -----SNYFAYNKLBPASGTTGLVNLGNTCYMNSALQCLVHPQLRDYFLYDGYDE 400
QY 771 LNRTPNIGMKGMKCYGLVQELW-----SGTQKVA--PLKLRWTIAKYAPRENGFQQ 824
DB 401 INEENPLGTHGVARAFSDLVQKLFQNRMSIQNRNAAPFPPSMFKSTIGHFNSMFSYMQ 460
QY 825 DSQELAFLLDGLHEDLNVRHEKPYE---LKDSDGRPDWEV---AAEAWNDHLNRNS 877
DB 461 DSQELAFLLDGLHEDLNRIIKEYTEKPSLSPGDDVNDVNVVKLADDTWEMHLKRNS 520
QY 878 IYVDLPHGQLRSQVKCKTCHISVRPDPNFIPLPMDSYHLEITVILKDGTPVRVG 937
DB 521 VIITDLFVGMKSTLYCPEQCNVSIPTDPYNDVTLLPLVDVTVDWTKIKIPPMN-SP 579
QY 938 LRLNMDKTYTGLKKQLSDCLGNSQILLAEVHGNIKNFQDNOKVRLSVSGFLCAFEI 997
DB 580 VELSKSTYMDLKNYVVKMSGLDPTLFCCEIFSNQI-----YVNYESTESNAQFLTQEL 635
QY 998 PVPVSPISASSPTQDTFSSSPSTNE-----MFTLTNGDL 1032
DB 636 IKFADDAVI-----FYELPVTNDNEVIVPVLNTRIEKGYKNAMLFGVPFFITLKEDEL 687
QY 1033 PRIFTPNGMPTV-----PCCTEKNTFTNGVNVGHMPSI-----PDSPTGY-- 1075
DB 688 NNPGAIRMKLQNRVHLSGGYIPFPEPVGNRTDFADA-----FPLLVEKYDVEFEQYKD 742
QY 1076 -----IIAHRKMMRTLYFLSSOKNRPISLFGMPLI----- 1106
DB 743 ILQYTSIKVTDKDKSPFSIKILSVK-----EQQFASNNRTGFN-FWTFISQLNLDKATD 796
QY 1107 VPCVTHTRKDLV-----DAWIOV-----SLASP-----LP 1134
DB 797 IDKLEVDVVKDIYVSSSLVDCAEGLVMQDDDEGTEGSEAKNFKSPFGSGDDEENKETVT 856
QY 1135 PQEASNAHQDCDDSMGY-----TDDVEDASTEPELTDKPEALDKIKDSLTSTPFAISM--N 912
DB 857 NNENVTNDRDEDMELTDDVEDASTEPELTDKPEALDKIKDSLTSTPFAISM--N 912
QY 1166 SCWCWYFRCGKIDCGEDRAPIGNAVIADWPTALHLRYQTSQERVVDDEHSVOS 1225
DB 913 DIIVCEW-----SELGSNEAFSDDK--IYNWENPA-----TLPNKL-ENAKLERS 955
QY 1226 RRAQAEPINLDSCLRAFTSEBELGENEMYCCKTHCLATKLDLWRLPPTLIHLKRF 1285
DB 956 -NAKERTIITDCLQLFSKPEILGLTDSWYCTKEHQATKIQIQLWNTPDILLHLKRF 1014
QY 1286 QVNGRWIKSQKIVFPRESFDPFSAFLVPRDPALCQHKPLTPQGBELSEPRILAREVKV 1345
DB 1015 ESQSPSDKIDATVNFPIITDLSRYVVKD----- 1045
QY 1346 DAQSSAGEEDVLLSKSPSSLSANI1SSPKGSSRSKSTGSCFSSKNKSNSSPRTLGHS 1405
DB 1046 -----DPRGL----- 1050
QY 1406 KGRRLPQIGSKNKLSSKENLDAKENGAGQICELADALSRGHVGGSQPELVTPQDHE 1465
DB 1051 ----- 1050
QY 1466 VALANGFLYEACNGYNGQIKNHSEEDSTDODREDTRIKPIYNLYAISCHSILGGG 1525
DB 1051 -----IYDLAYVDNHYGGLGG 1067
QY 1526 HVTYAKN-PNCKWYCYNDSCKELHPDEIDTDSAVILFY 1564
DB 1068 HTAYVKNFADKNYTFDDSRVETAPENSIAGSAYILFY 1107

RESULT 10

UBP1 SCHPO STANDARD; PRT; 849 AA.

AC Q9USM5.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable ubiquitin carboxyl-terminal hydrolase 1 (EC 3.1.2.15)

DE (Ubiquitin thiolesterase 1) (Ubiquitin-specific processing protease 1)

DE (Deubiquitinating enzyme 1).

GN UBPI OR SPCC16A11.12C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Cellins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gattrell S., Hamlin N., Harris D., Hidaigo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,

RA Weijens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Gadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet W., Gallard J., Tallada V.A., Garzon A., Rhode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe."

RL Nature 415:871-880(2002).

CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =

CC ubiquitin + a thiol.

CC -!- SIMILARITY: Belongs to peptidase family C19.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AL109957; CAB53084.1; --

DR F01; T41085; T41085. --

DR GendB_Sfombe; SPCC16A11.12c; --

DR InterPro; IPR006615; DUSP.

DR InterPro; IPR001394; UCH-2.

DR Pfam; PF00443; UCH; 1.

DR SMART; SM00695; DUSP; 1.

DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS00973; UCH_2_2; 1.

DR PROSITE; PS02335; UCH_2_3; 1.

DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.

FT ACT_SITE 288 BY SIMILARITY.

FT ACT_SITE 798 BY SIMILARITY.

FT ACT_SITE 806 BY SIMILARITY.

SQ SEQUENCE 849 AA; 98655 MW; C8118042229BA9CB CRC64;

Query Match 8.1%; Score 693.5; DB 1; Length 849;
 Best Local Similarity 21.3%; Pred. No. 5.6e-36;
 Matches 233; Conservative 160; Mismatches 321; Indels 379; Gaps 32;

QY 534 PVKATSLTLEGRLKRTQQLNGRDYEMVPPVWRALYHYWG-ANLALPRPV--IKNSKT 590
 DB 70 PITQWRLLDEKNEKHS--LEESIDYSVASLWMLVWVFWGLEGLAERKVLVGLAAE 127

QY 591 DIPELELP-RVLLFLRQOPATRTQOSNIWNMGNV-SP-----NAPLKRVLAYTGCFSR 644
 DB 128 QKPFVDIYPIINFTHLVFDP-----INGENTSPLYQIDEPYHSDEFAVAFSPSR 177

QY 645 MOTIKIEIHYLSQRRIKE-EDMRLWLYNSEN-----YLTLLDDEHKL-----EY----- 689
 DB 178 SDTLRSLYKQVMEAFQISDGTSLRWLYLNKSNLSRFSVLSSEFNDDQPAIALLSAYAVCM 237

QY 690 ---LKIQDEQHLVIEVRKNDKSWPEMSFIANSKIDRHKVPTKGTGATGLNLTGTCFVN 746
 DB 238 IFEIDIADGS-LLELFQHPNGEWLSD-----SITKEQLTINKEIGLCGLYNLGNSCYMN 291

QY 747 SSIQCVSNTOPLTOYFISGRHLYELNRNTPIGMKHMAKCYGDLVOEL-WSGTQKNVAPL 805
 DB 292 SALQCMHITHETKFLSDSYEKDINYNPLGWMGKVALSYASLLKMIHTADHMSVSPS 351

QY 806 KLRWTIAYAPRNFQOQDSOELLAFLLDGLHDLNRVHEKPYVELKDSGRPD----- 860
 DB 352 SPKFIIGFNTYFSGYRQDSQEFIAFFLDGLHDLNRVHEKPYVELKDSGRPD-----RPLDFDEH 405

QY 861 ----NEVAAEADNHLRNRSTVVDLPHGLQSRVOKVCKTCGHSVRDPENFLSLPLMD 916
 DB 406 PHVORVANQCWIITKNDLSIIVLFQOMKYSTLECSIQKSTAFDFPFVLTPLTPTS 465

QY 917 SYMHLEITVIKLDGT-TPVRYGLRLNMDKTYGLKKQLSD---LGLNSQEQLLAEVHGS 972
 DB 466 AKWRHKVAVVPPFGTQSPVELYLLLMESTVQMFKQATEKLQKMGLEGELTACDIYRG 525

QY 973 NIKNPPQDNQKRLSVSGFLCAFEFVPPVSPISASSTPTQDFSSSPSTNEMFTLTNGDL 1032
 DB 526 KYVKLVKNDKISKKI-----

QY 1033 PRPIFIPNGMPTVPCGTGKNTFNGMVG-HMPSLPDPSPTGYIIAHRKMRTELAYFL 1091
 DB 542 -----HKWDHVLVYGSTANGLTPIVHGCKRPAMPG-----

QY 1092 SQKQKPSLFGMPLIVPCTVTRKKD-----LYDAVWIOVSRSLASPLPPQ 1136
 DB 573 SYQSN--DVFGFPL---QLNVRSRNLTNDLVKEIVELYRYAGIDVAIGTLQGLKRM 626

QY 1137 EASNEAQDCCDSMGYQYPTLRVWQDGNSCAWCPWYFCRGCKIDCGEDRAFIAGNAYIA 1196
 DB 627 ESKAGWECIKIEVK---RFEIVEE-----EIV 653

QY 1197 VMDPTALHRLYQTSQ-ERVVDEHESVQSRAQAPINLDSCLRAFTSBEELGENEMY 1255
 DB 654 ID-DKTVIMCLMNDQYKELFYNCWEIPEKIQFMESITLEDCLLEFSPKPEQLDQDSWY 712

QY 1256 CSKCKTHCLATKLDLRLPILIIHLAKRFQVNG---RWIKSQKIVKFPRESPPSAPFL 1312
 DB 713 CPGCKAFRPRATKRLRWLPKLVILHNRPSGHGDLRRRKRRLRVVYVPFDLNLKQFL 772

QY 1313 VPRDPALCQHKPLTPQGDSELPRLAREVKVDAQSAGEEDVLLSKSPSLSANISS 1372
 DB 773 SP-----

QY 1373 PKGSPSSSRKSGTSCFSSKNSSPSPRTLGRSKGRRLPQIGSKNKLSSSKENLDASKE 1432
 DB 775 -----

QY 1433 NGAGQICELADALSRGHVLGSGQPELVTPQDHEVALANGFLYEACGNGYSGOLGNHS 1492
 DB 775 -----FIKDHEWLSQKS-----

QY 1493 EBDSTDQREDTRIKPIYINLYAISCHSILGGHVVYAKNPCK-WYCYNDSSCKELHP 1551

OC Gallus.
 RN NCBI_TaxID=9031;
 RP [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=97467343; PubMed=9325273;
 RA Baek S., Choi K.S., Yoo Y.J., Cho J.M., Baker R.T., Tanaka K.,
 RA Chung C.H.;
 RT "Molecular cloning of a novel ubiquitin-specific protease, UBPA1, with
 RT isopeptidase activity in chick skeletal muscle.";
 RL J. Biol. Chem. 272:25560-25565(1997).
 RL -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF016107; AAC13729.1; -
 DR MEROPS; C19.013; -
 DR InterPro; IPR001394; UCH-2.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS02335; UCH_2_3; 1.
 KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 28 28 BY SIMILARITY.
 FT ACT_SITE 301 301 BY SIMILARITY.
 FT ACT_SITE 309 309 BY SIMILARITY.
 SQ SEQUENCE 357 AA; 40931 MW; EC3B6454937C455 CRC64;
 Query Match 5.2%; Score 446; DB 1; Length 357;
 Best Local Similarity 17.9%; Pred. No. 6.7e-21;
 Matches 152; Conservative 51; Mismatches 134; Indels 512; Gaps 9;
 QY 719 SSKIDRHVPFEKATGLSNLGNFCFNNSSIQCVSNQPLTQYFISGRHLYELNRNTPIG 778
 DB 11 SSKV-----VOGLTGLRLNLTFCFNNLSLQCLNLTKELDYCLNQYLRLDN--NNSR 61
 QY 779 MKGHMAKCYGLVQELW-SGTQKNVAPLKLRTWTAKYAPRNGPQQDSQELALFLDGL 837
 DB 62 MRTLMEFAKLIQLLWTSNDVSPSEFKTQRYAPRFVGNQDAQEFRLFLDGL 121
 QY 838 HEDLNRYHEKPYVELKSDGRPDWEVAAEAWNDHLNRNRISVVDFHGLQRSOVKCKTCG 897
 DB 122 HGEVNRVLRPRANATLDLHPDDEKSRQWRVQEREDSRVSDLLFVGQLKSLTCEG 181
 QY 898 HISVRFPDFNLSPLPMDSNMHLIEITVILKDGTPVRYGLRLNMDKTYGLKKQLSDLC 957
 DB 182 YCSAFDFPFDLSLPIPKGY----- 202
 QY 958 GLNSEQILLAEVHSGNIKNFPQDNQKVLRSVSGFLCAFEIPVPVSPISASSPTQDFSSS 1017
 DB 203 ----- 202
 QY 1018 PSTNEMTLTNGDLPRPIFIPNGMNTVVPVCGTEKNTFMVNGHMPSLDPSFTGYII 1077
 DB 203 ----- 202
 QY 1078 AVHRKMRTELYFLSSQKNRPSLFGMPLIYVCTVHTRKDLVDVAVIQVSRLASPLPQB 1137
 DB 203 ----- 202
 QY 1138 ASNHAQCDSDSMGYQYPTFLRVQKDGNSCAWCPWYRCRGCKIDCEDRAFTGNVAV 1197
 DB 203 -----GE----- 204
 QY 1198 DWDFTALHRYQTSQERVDEHESVEOSRAQAEPINLDSCLRAFTSEBELGENMYICS 1257

DB 205 -----VTLMDCURLFTKEDVLDGDKPTCC 229
 QY 1258 KCKTHCLATFKKLDLWRLPILIIHLKRFQFVNGRWIKSQKIVKFRFESFDFPSAFLVPRDP 1317
 DB 230 RCKARTRCTKFSIQKPKILVLHLKRFSEARIRASKLTTTFVNFPLKOLDLREF----- 283
 QY 1318 ALCOHKPLTPQDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSSLSANIISSPKSP 1377
 DB 284 ----- 283
 QY 1378 SSSRKSGTSCPSSKNSPNSPRTIGRSKGRLLRPLQIGSKNKLSSKENLDASKENGAGQ 1437
 DB 284 -----ASQ 286
 QY 1438 ICELADALSRGHVLSGQPELVTPQDHEVALANGFLYEHEACNGYNGQLGNHSEEDST 1497
 DB 287 SC-----NHA----- 291
 QY 1498 DQREDRTKIPYLNLYAISCHGILGGHYVYAKNP-NCKWYCYNDSSCKELHDEIDT 1556
 DB 292 -----VYNLYAVNSHSGTTMGHYTAYCKSPISSEHSHFNDSRVTPMSSSHVRS 340
 QY 1557 DSAYILFYE 1565
 DB 341 SDAYLLFYE 349
 RESULT 14
 UBPA1_HUMAN STANDARD; PRT; 605 AA.
 AC 075604; Q9BQ21;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
 DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
 GN USP2 OR UBPA1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RA Gong L., Yeh E.T.H.;
 RT "Cloning and expression of the human and mouse UBPA1";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.M., Hsieh F.,
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiy S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Maman A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

CC NCBI_TaxID=10090;
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Gong L., Yeh E.T.H.;
 RT "Cloning and expression of the human and mouse UBPA1";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AF079565; AAC28393.1; -;
 CC MEROPS; C19.013; -;
 CC MGD; MGI:1858178; Usp2.
 CC InterPro; IPR001394; UCH-2.
 CC Pfam; PF00443; UCH; 1.
 CC PROSITE; PS00972; UCH_2_1; 1.
 CC PROSITE; PS00973; UCH_2_2; 1.
 CC PROSITE; PS02035; UCH_2_3; 1.
 CC Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 24 24 BY SIMILARITY.
 FT ACT_SITE 297 297 BY SIMILARITY.
 FT ACT_SITE 305 305 BY SIMILARITY.
 SQ SEQUENCE 353 AA; 40581 MW; 4FFB39A225FE8F11 CRC64;
 CC -----
 CC
 CC Query Match 4.8%; Score 412; DB 1; Length 353;
 CC Best Local Similarity 16.9%; Pred. No. 9.3e-19;
 CC Matches 144; Conservative 59; Mismatches 134; Indels 516; Gaps 10;
 QY 717 ANSSKIDRHKVPTEKGATGLNLTGNTCEPMNSIQCVSNTPQLTQYFISGRHLYELNRTNP 776
 Db 5 AKNSK-----SAQGLAGRLNLTGNTCEPMNSIQCVSNTPQLTQYFISGRHLYELNRTNP 56
 QY 777 IGMKGHMA--KCYGLVQELMSGTKN-VAPLKLRTWTAKYAPRFGNQDQSOELLAF 833
 Db 57 ---SAHTALMEEFAKLIGITWTSNDVVSSEFKTQIQRYAPRFGNQQDAQEFLRPL 113
 QY 834 LDGLHEDLNVRHEKPYVELKDSGRPDWEVAEAWDNHLRNRSIVVDLPHGLRSQVKC 893
 Db 114 LDGLHNEVNRVAARPKASPELTLHLDEKGRQWRKYLERSRIGDLFVGQLKSLTLC 173
 QY 894 KTCGHISVRFPDPNPLSLPLPMDSTMHLEITVTKLDGTPVRYGLRLNMDEKYLGLKKQL 953
 Db 174 TDCGYCSTVDFDFWDLSPIAKGY----- 198
 QY 954 SDLCGLNSEQILLAEVHGSNIKFPDQNKVRLSVSGFLCAFEIPFVSPISASSPTQTD 1013
 Db 199 ----- 198
 QY 1014 FSSSPSTNEMTLTNGDLPRPIPIPGMNTVVPCTGTEKNTGNGVNGHMPSLDPSPT 1073
 Db 199 ----- 198
 QY 1074 GYIIAVHKKMRTLYFLSSQKNRPSLFGMPLIVPCTVHTRKKDLYDAVMIQVSRLASPL 1133
 Db 199 ----- 198
 QY 1134 PQEASNEAQDCDDSMGYQYPTLRRVQKGNCSAWCPWYRCGCKIDCGEDRAFTGNA 1193
 Db 199 -PE----- 200
 QY 1194 YIADVWDPTALHLRYQTQSERVVDHESVEQSRRAQEPINLDSCLRAFTSERELGENEM 1253
 Db 201 -----VTLMDCMRLFTKEDILDGDEK 221

QY 1254 YYCSKCKTHCLATKLDLWELPILIIHLKRQFVNGRWIKSQXIVKFPRESFDPSPALV 1313
 Db 222 PTCRCRAKRCIKFESVQRFKILVHLKRFESRIRTSKLTTFVNFPLRDLID----- 275
 QY 1314 PRDPALCOHKPLTPQGDSEPRILAREVKVDAOSSAGEEDVLLSKSPSSLSANIISP 1373
 Db 276 ----- 275
 QY 1374 KGSFSSSRKSGTSCFSSKNSPNSPRTLGRSKGRILRLPOIGSKNKLSSKENLDASKEN 1433
 Db 276 -----LEEFASENT----- 284
 QY 1434 GAQICELADALSRGHVLSGSPQLVTPQDHEVALANGFLYEHEACNGYNSGQLGNHSE 1493
 Db 285 -----NHA- 287
 QY 1494 EDSTDQREDTRIKPIYNIYALISCHSGLIGGHYTYAKNP-NCKWYCVNDSCKELHPD 1552
 Db 288 -----VYNLYAVSNHSGTTMGGHYAYCRSPVTGWHTFNDSVTPMSSS 332
 QY 1553 EIDTDSAYILFYE 1565
 Db 333 QVRTSDAYILFYE 345

Search completed: February 11, 2004, 15:39:17
 Job time : 32 secs